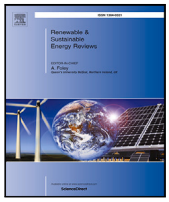
Renewable and Sustainable Energy Reviews 154 (2022) 111768

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser



Solar forecasting with hourly updated numerical weather prediction

Gang Zhang [a], Dazhi Yang [a,*], George Galanis [b], Emmanouil Androulakis [b]

[a] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China
[b] Section of Mathematics, Mathematical Modeling and Applications Laboratory, Hellenic Naval Academy, Hatzikiriakion, Piraeus, Greece

ARTICLE INFO

Keywords:
Numerical weather prediction
Hourly solar forecasting
Post-processing
Kalman filtering

ABSTRACT

Solar forecasters have hitherto been restricting the application of numerical weather prediction (NWP) to day-ahead forecasting scenarios. With the hourly updated NWP models, such as the National Oceanic and Atmospheric Administration's Rapid Refresh (RAP) or High-Resolution Rapid Refresh (HRRR), it is theoretically possible to utilize NWP for hour-ahead solar forecasting. Nonetheless, for NWP-based hourly forecasts to be useful, they ought to be post-processed, because such forecasts are almost always affected by the inherent bias of dynamical weather models. In this regard, Kalman filtering is herein used as a post-processing tool. Using one year of RAP and HRRR forecasts and ground-based observations made at seven geographically diverse locations in the contiguous United States, it is demonstrated that the post-processed versions of hourly updated NWP forecasts are sometimes able to attain a higher accuracy than those from classic time-series families of models, not only at long-range, but also at short-range forecast horizons. Although these improvements are marginal in terms of squared loss (about 1%–2%), since NWP models have a very different forecast-generating mechanism (solving the governing equations of motion) from that of time series methods (extrapolating data), NWP-based forecasts can be expected to be less correlated with those forecasts from time series models. Consequently, one should find this diversity profoundly rewarding during forecast combination, for the combined forecasts are able to consistently result in smaller error than the best component forecasts.

1. Introduction

The two topics, namely, solar irradiance forecasting and solar power forecasting, are jointly known as *solar forecasting*. In fact, it is not possible to isolate one topic from another, so long as goodness of forecasts is to be valued at any rate. Solar irradiance, as the chief weather variable influencing the power production of solar energy systems of various sorts, has a decisive effect on the accuracy of solar power forecasts. If a forecaster wishes to obtain high-quality solar power forecasts, the ability to generate and use irradiance forecasts is absolutely necessary [1]. Stated differently, the best-possible photovoltaic (PV) or concentrating solar power (CSP) forecasting procedure always starts from generating irradiance forecasts, and then converts those irradiance forecasts to power forecasts via a model chain [2,3]. On the other hand, an irradiance forecast, by itself, has no intrinsic value [4], instead, it is its ability to influence decision-making that assigns a value to the forecast [5].

Solar forecasting is becoming an essential element of modern power system operations, which has thus led to the recent wave of rapid development of theories and principles related to solar forecasting [6]. With an increasing penetration of solar power into the electricity grid, traditional power system operations, such as load flow, day-ahead and intra-day unit commitment, economic dispatch, or real-time regulation, are faced with new challenges pertaining to the inherent variability and uncertainty of solar power [7,8]. For instance, in traditional day-ahead unit commitment, since thermal generators are dispatchable, the only source of uncertainty in the optimized generation schedule comes from the electric load, which can be forecast fairly accurately to an error size of 3% [9]. This 3% error can subsequently be absorbed during intra-day unit commitment and real-time regulation. In contrast, high penetration of solar power would cause severe deviations in unit-commitment result from the actual operational scenario, because the forecasts for net load (load minus solar) are inaccurate—day-ahead solar forecast error is about 25% [10], which propagates into large net load forecast error. An imprecise day-ahead generation schedule puts much pressure on the subsequent intra-day and real-time operations, which eventually translates into higher reserves required, and thus higher cost. On this

Abbreviations: ECMWF, European Centre for Medium-Range Weather Forecasts; HRRR, High-Resolution Rapid Refresh; KGC, Köppen–Geiger climate classification; NAM, North American Mesoscale; NOAA, National Oceanic and Atmospheric Administration; NWP, numerical weather prediction; MOS, model output statistics; QC, quality control; RAP, Rapid Refresh; SURFRAD, Surface Radiation Budget Network; UTC, Coordinated Universal Time

* Corresponding author.
E-mail address: yangdazhi.nus@gmail.com (D. Yang).

https://doi.org/10.1016/j.rser.2021.111768
Received 26 May 2021; Received in revised form 1 September 2021; Accepted 12 October 2021
Available online 2 November 2021

point, if the quality of solar forecasts can be improved, a higher value can be materialized from having a reliable and stable power system with minimal reserves.

Solar forecasters, who are primarily concerned with predicting the future values of solar irradiance as a means to forecast solar power, have been firmly believing in two things over the past decade. One of those is that solar forecasting, of which the accuracy is most heavily impacted by moving clouds, can benefit substantially from those methods that aim to capture cloud dynamics by physical means, such as sky camera, satellite, and numerical weather prediction (NWP) [11]. The second belief is that the physics-based forecasting methods, due to the inherent limitation of the precision of analysis and the incomplete understanding on the atmospheric processes, often require statistical or machine-learning-based post-processing [12]. Combining these two views, one can readily conclude that physics-based solar forecasting with appropriate data-driven post-processing is what solar forecasters believe to be the best practice, insofar as the present state of affairs is concerned [13,14].

Regarding the physics-based solar forecasting methods, people have hitherto been associating exogenous data of various kinds with forecast horizons on which they are thought most useful [15]. More specifically, anyone who has read any general review on this topic would have undoubtedly learned that sky camera, satellite, and NWP output data are most functional for intra-hour, intra-day, and day-ahead solar forecasting, respectively. In fact, such association has been strengthened repeatedly by the expanding and propagating literature—the review by Inman et al. [11], which advocates the method–horizon association, is by far one of the highest cited reviews on solar forecasting. However, the present authors no longer believe in such association being appropriate.

We should wish to make our reasons for this rejection clear as we proceed: (1) largely owed to the advent of the latest generation of meteorological satellites and the increase in computation power, both the characteristic scales of remote-sensing data and the running cycles of NWP models have reduced dramatically in the past few years, with the latest satellite-derived irradiance reaching a 5-min–2-km native resolution [16] and the operational NWP being updated hourly [17], allowing forecasting at higher resolutions and/or for shorter horizons to be performed leveraging these data; and (2) as solar forecasters are being exposed to an ever-greater number of advanced data-fusion tools, integrating forecasts of different temporal resolutions is now possible [18–20]. In other words, both the quality of the data sources and the forecasting techniques have improved over the years; they present solar forecasters with opportunities to tackle tasks that have been perceived previously as unfeasible. In this article, one task of that sort shall be investigated: *performing hour-ahead solar forecasting via NWP*.

2. Backgrounds on hour-by-hour NWP

One of the great leaps in scientific theories related to weather forecasting occurred in the early 20th century, when the Norwegian scientist Vilhelm Bjerknes presented his vision on the so-called "rational solution of forecasting" in 1904 [21]. Bjerknes visioned that a forecasting task ought to be split into two steps, one diagnosis and one prognosis. The diagnosis step (or *analysis*) gathers information in regard to the initial state of the atmosphere, whereas the prognosis step (or *forecast*) develops future states out of the initial one based on some physical principles called dynamics. Two decades after that, Bjerknes' idea was precisely implemented by Lewis Fry Richardson [22], who presented the first-ever NWP "run" for a single day—20 May, 1910—over central Europe with just a handful of grid cells. Nonetheless, the computation took 6 weeks which, though seems less impressive given the present technology, is in fact admirable considering the fact that everything was calculated manually. Indeed, the atmosphere is by far the most complex dynamical system that men have dealt with, so complex that it requires hours to complete a single run using modern supercomputers, which are often available just at national weather centers and space agencies [23]. For that reason, NWP models are typically run only a few times a day, by a countable few groups of people (pers. comm. Jan Kleissl). For instance, in America, the North American Mesoscale Forecast System (NAM) issues forecasts four times a day at 00, 06, 12, and 18 UTC, the same goes for the European Centre for Medium-Range Weather Forecasts (ECMWF) in Europe.

That said, there are certain specialized fields, such as aviation, severe weather watches and warnings, hydrology, or energy, which require an forecast update frequency higher than 6 hourly. In concert with the increasing use of automated decision-making software and tools, the capability of issuing frequently updated NWP forecasts is becoming progressively more desirable. To that end, the National Oceanic and Atmospheric Administration (NOAA) introduced the first-ever hourly updated operational NWP system—the Rapid Update Cycle (RUC)—in 1998 [24]. By assimilating observations in an hour-by-hour fashion, RUC served as an initial U.S. weather situational-awareness model [17]. Over the years, RUC had evolved several times and was eventually replaced by the Rapid Refresh (RAP) model in 2012, which is the model of interest in this article. Similar to any operational forecasting system, RAP has also been upgraded numerous times, in 2014 (v2), 2016 (v3), 2018 (v4), and 2020 (v5).[1] For technical specifications in regard to RAP, such as data assimilation, model component, or lower-boundary treatment, the reader is referred to Benjamin et al. [17].

It is noted here that a dynamical weather model is based on a spatial discretization (or grid) of the differential equations of motion. Whereas the so-called "dynamical core" describes the resolved scales of motion, physical parameterizations provide estimates of the grid-scale effect of processes on the atmospheric state that cannot be resolved. In other words, NWP models cannot resolve sub-grid scale weather features, which refers to those processes evolve within a grid cell, such as the turbulent eddies due to surface structures like buildings or tall trees. In this regard, the finer the spatial discretization of an NWP model is designed to be, the stronger its resolving power is expected to be. (It should be noted however that there are limitations in this positive correlation, since from a resource viewpoint, it is uncertain whether the obtained improvements are able to compensate for the costly computation that is required to generate those forecasts [25].) The horizontal grid spacing of RAP is 13 km, which is comparable to the 12-km NAM model that is popularly used in solar forecasting [26–28]. Notwithstanding, there is another model, with an even higher resolution, in the RAP family, that is, the High-Resolution Rapid Refresh (HRRR), which is also hourly updated and has a 3-km horizontal resolution. HRRR is nested within the RAP domain, in that, it is only available for the lower 48 United States, whereas RAP covers greater North America and the Arctic. Currently, HRRR is at version 4, which was implemented at 12 UTC on December 2, 2020, together with RAPv5.[2]

Both RAP and HRRR are updated hourly, but their forecast horizons are slightly different, see [29]. RAPv4, on one hand, issues 0–21-h-ahead forecasts for all runs except for those starting at 03, 09, 15, and 21 UTC, during which 0–39-h-ahead forecasts are issued. RAPv5 extends the forecasts for those four runs further to 51 h ahead. HRRRv3, on the other hand, issues 0–18-h-ahead forecasts for all runs except for those starting at 00, 06, 12, and 18 UTC, during which 0–36-h-ahead forecasts are issued. HRRRv4 extends the forecasts for those four runs further to 48 h ahead. The frequent updates of RAP and HRRR, in contrast to NAM or ECMWF which issues new forecasts only every six hours, allow one to make new inquiries regarding the optimal solar forecasting practices, which shall be elaborated next.

[1] See details at https://rapidrefresh.noaa.gov/.
[2] https://rapidrefresh.noaa.gov/hrrr/.

3. Research questions

A total of four questions are raised and addressed in this article. Two of them are related to verification, which is absolutely vital, because any forecast user would need to have a basic level of understanding on the quality of the received forecasts. The other two are those of post-processing, which affects the "final" version of forecasts that is used for decision-making.

- At what rate does the performance of hourly updated NWP deteriorate?
- How does the hourly updated NWP compare to univariate statistical and machine-learning models in terms of forecast quality?
- Is Kalman filtering an adequate tool for post-processing hourly updated NWP?
- Is the post-processed hourly updated NWP significantly better than the post-processed day-ahead NWP?

The theoretical background and practical concerns regarding these four questions are detailed in the following four subsections.

3.1. Performance deterioration and predictability

Virtually all forecasters have the tendency to take as true that, with everything else being equal, short horizons are more predictable than the long ones. Diebold and Kilian [30] argued that *predictability* should be defined as one minus the ratio of expected losses of an optimal short-run forecast and an optimal long-run forecast, that is:

$$\text{predictability} = 1 - \frac{\mathbb{E}\,(f^s - x)^2}{\mathbb{E}\left(f^l - x\right)^2}, \tag{1}$$

where x denotes observation (or verification), f^s and f^l denote the optimal short- and long-run forecasts, respectively, and $\mathbb{E}$ is the expectation operator. Here, the squared loss is used without loss of generality.[3] This definition of predictability is, however, practically challenged, for one is unable to know what constitutes an optimal forecast. Meteorologists often assume there exists a perfect model, which is able to describe the laws of motion exactly. Similarly, statisticians often assume there exists a data-generating process, based on which the observations materialize. Both schools of thought are but theoretical, and are unable to be verified. Notwithstanding, the formulation in Eq. (1) does allow one to quantify the rate at which the performance of any forecasting system deteriorates.

For the two NWP models at hand, the longest hourly updated forecast has a horizon of $h = 21$ for RAP and $h = 18$ for HRRR. Therefore, one can compute the ratio of expected losses—using forecasts at each shorter horizon with respect to those at the longest horizons—21 or 18 times, depending on the model. Mathematically,

$$\text{IRA}(h, h_{\max}) = 1 - \frac{\mathbb{E}\left(f^h - x\right)^2}{\mathbb{E}\left(f^{h_{\max}} - x\right)^2}, \tag{2}$$

where $h = 1, 2, \ldots, 21$ (or 18), f_h corresponds to a h-step ahead forecast, and $f^{h_{\max}}$ is the forecast at the furthest horizon. It should be noted that Eq. (2), though has the same function form as Eq. (1), shall be interpreted in a different way. Since the ratio of expected losses is calculated based on forecasts from the same model, but at different horizons, Eq. (2) can be viewed as a measure of *internal relative accuracy* (IRA) of the model itself, in that, the IRA is low if $\mathbb{E}(f^h - x)^2 \approx \mathbb{E}(f^{h_{\max}} - x)^2$, and is high if $\mathbb{E}(f^h - x)^2 \ll \mathbb{E}(f^{h_{\max}} - x)^2$. It is also obvious that IRA should drop as forecast horizon increases, and it eventually reaches 0 when the forecast horizon is at the maximum.

That said, it would not be unreasonable to take $\text{IRA}(1, h_{\max})$ as a proxy of predictability; the reader is referred to Yang et al. [32] for a full account of this proposition. For locations with $\text{IRA}(1, h_{\max}) \approx 0$, which implies that the forecast accuracies at shortest and longest horizons are nearly identical, the predictability must be said to be low, for there is not much to do besides issuing the long-run climatology forecasts.[4] In contrast, for locations with a significantly high $\text{IRA}(1, h_{\max})$, the predictability is said to be high, because the accuracies at shortest and longest horizons are distinguishable. Under such scenarios, one has grounds to believe that there are plenty of actions that can be taken, at short horizons, on top of issuing just the long-run climatology forecasts. On this point, the first mission of this article is to investigate how IRA of RAP and HRRR varies under different climate conditions, and thus to understand, based on empirical evidence, how predictability of solar irradiance may behave.

3.2. Comparative forecast verification

According to Murphy [33], all forecast verification methods must be one of either absolute verification or comparative verification. In absolute verification, forecasters evaluate the performance of a single forecasting system. During the process, the three aspects of goodness of a forecast, namely, consistency, quality, and value [see 5, for details], can be accounted for. Absolute verification, however, offers no freedom to forecast users, insofar as the question "Can we do better?" is concerned. It thus implies that absolute verification has little practical relevance, for forecast users have no other choice but can only take whatever that is being offered as acceptable. On the other hand, comparative verification compares two or more forecasting systems, with the goal of selecting one which has the highest performance and/or value, such that it can be used for subsequent decision-making. In general, comparative verification can be conducted across different forecasting situations, i.e., the sets of observations that correspond to forecasts generated by various forecasting systems are different. The trick to do this is by involving a naïve reference model, such that forecasts of different systems are gauged in terms of their relative performance to their respective reference forecasts—the resultant measure of performance is known as the *skill score*.

Commonly used naïve reference methods include persistence, which uses the most recent observation as forecasts, and climatology, which issues the long-term mean as forecasts. When multiple naïve reference methods are present, the one with the highest accuracy should be used as the standard of reference. Since persistence almost surely outperforms climatology at short horizons and the situation reverses at long horizons, a group of 33 solar forecasting experts have been advocating the use of an optimal convex combination of climatology and persistence [4]. The technical details of the climatology-persistence combination can be found in [34,35]. One should note that, in order to account for the double seasonality (yearly and diurnal cycles) in solar irradiance and solar power, and to avoid exaggerating the skill score, all reference methods ought to be applied on the deseasonalized quantity called *clear-sky index*, which is the ratio between the irradiance (or PV power) and its clear-sky expectation [4,34,35].

Besides examining the skill scores of RAP and HRRR, which would be the second mission of this article, it is also necessary to compare their performance with slightly more advanced forecasting methods, such as those time series and machine-learning methods, which have been regarded as adequate for hourly forecasting [36,37]. In the literature, information on this sort of comparison is limited, for NWP forecasts are mostly used in day-ahead settings, where pure statistical

[3] Another scoring function, namely, the absolute error is also adequate. The choice between absolute and squared errors, however, depends on the directive under which the forecasts are requested. The choice of scoring function should follow the forecast directive, for a model optimization performed on a scoring function other than the one required is unlikely to be truly optimal [31].

[4] In meteorology, forecasts at the longest horizon often simply reduce to climatology.

and machine-learning methods usually cannot be extrapolated that far with decent accuracy. Similarly, it would be equally unfair to compare day-ahead NWP with hourly data-driven models, for sequential information update is critical for short horizons, and day-ahead NWP only refreshes its information once a day. With the hourly updated NWP, one has now the basis to conduct such a comparison, and consequently, the results would advise whether or not NWP can be justified for short-term solar forecasting.

3.3. Sequential versus batch post-processing

Model output statistics (MOS) has been the most popular post-processing technique for deterministic NWP-based solar forecasts [12]. The distinction one should however make is between MOS and ensemble model output statistics (EMOS), of which the latter handles the calibration of ensemble forecasts. This article shall only restrict its discussion to MOS; the reader is referred to Yang and van der Meer [12] for a review of EMOS and other probabilistic post-processing techniques.

The first major work on solar forecasting that introduced MOS to the field is one by Lorenz et al. [38], who used a forth-degree polynomial regression to model the bias of NWP forecasts:

$$\begin{aligned}\text{bias}_t = {} & \beta_1 + \beta_2\hat{\kappa}_t + \beta_3\hat{\kappa}_t^2 + \beta_4\hat{\kappa}_t^3 + \beta_5\hat{\kappa}_t^4 + \beta_6\cos\theta_t + \beta_7\cos^2\theta_t \\ & + \beta_8\cos^3\theta_t + \beta_9\cos^4\theta_t + \varepsilon_t, \end{aligned} \tag{3}$$

where $\hat{\kappa}_t$ and θ_t are NWP-forecast clear-sky index and zenith angle at time t, respectively; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_9)^\top$ are regression coefficients to be estimated; and ε_t is an error term for the regression. Clearly, before this model can be of use, $\boldsymbol{\beta}$ needs to be estimated through some training data. It also implies that once $\hat{\boldsymbol{\beta}}$, i.e., the vector of fitted regression coefficients, is obtained, one can post-process the new forecasts in batches, so long as $\hat{\kappa}_t$'s are available. Although there have been works demonstrating fitting MOS coefficients by season or by other data-stratification variables [39], this practice must not be regarded as general, for most works in the literature treat the MOS coefficients as static.

In contrast, Kalman filtering, which is the post-processing technique adopted in this work, is a sequential procedure. Based on a sequence of forecasts, and when ground-truth becomes progressively available as time passes, the sequence of forecasts can be filtered. Whereas the Kalman filtering procedure is reviewed thoroughly in Section 6.1, its modeling principle is briefly introduced here. A classic Kalman filter is a dynamical linear model, which has the form:

$$y_t = \alpha_{1,t} + \alpha_{2,t}x_{1,t} + \alpha_{3,t}x_{2,t} + \cdots + \alpha_{m,t}x_{m-1,t} + \varepsilon_t, \tag{4}$$

where y_t is the variable to be modeled; $x_{i,t}$ with $i = 1, \dots, m-1$ are $(m-1)$-dimensional inputs; $\boldsymbol{\alpha}_t = (\alpha_{1,t}, \dots, \alpha_{m,t})^\top$ is a length-m vector of coefficients; and ε_t is an error term. As indicated by the subscript t, all quantities in Eq. (4) are allowed to be time-varying. Suppose one adopts the same inputs as those used in Eq. (3), the Kalman filter model would be identical to Lorenz's MOS in terms of form:

$$\begin{aligned} y_t = {} & \alpha_{1,t} + \alpha_{2,t}\hat{\kappa}_t + \alpha_{3,t}\hat{\kappa}_t^2 + \alpha_{4,t}\hat{\kappa}_t^3 + \alpha_{5,t}\hat{\kappa}_t^4 + \alpha_{6,t}\cos\theta_t + \alpha_{7,t}\cos^2\theta_t \\ & + \alpha_{8,t}\cos^3\theta_t + \alpha_{9,t}\cos^4\theta_t + \varepsilon_t. \end{aligned} \tag{5}$$

The answer to the question "whether or not sequential/dynamic post-processing is more advantageous than batch/static post-processing", like many others, depends on perspective. Sequential/dynamic post-processing, on one hand, tends to provide better results, for recent observations carry more relevant information than that consisted in distant ones. In other words, sequential/dynamic post-processing is always conducted with reference to the most recent observations. Therefore, in this viewpoint, Kalman filtering is, in principle, expected to be better than MOS. On the other hand, sequential/dynamic post-processing is limited in terms of forecast availability. In order to perform filtering, both the current observation and the h-step-ahead forecast need to be available. This is, however, problematic when the forecasts are not issued in an hour-by-hour fashion—see Sections 3.4 and 6.1.3 below. In short, Kalman filtering might not be as an efficient method as it would be had the forecasts been generated with most recent information [20]. MOS is not subjected to this pitfall, because it is a batch technique which, after the model is fitted, does not require observations to operate; it is, hence, suitable for post-processing day-ahead forecasts. For more information on inadequacy of Kalman filtering in day-ahead solar forecasting, and various ambiguities of this sort found in the literature, the reader is referred to Yang [40]. In conclusion, the third mission of this article is to examine the amount of improvements on NWP forecast accuracy that can be achieved through the means of Kalman filtering.

3.4. Filtering for daily, 6-hourly, and hourly updated NWPs

In the literature of solar forecasting, there are works that have attempted to post-process NWP forecasts using Kalman filters, such that the results can be regarded as hour-ahead forecasts [e.g., 41,42]. The general practice is, however, to perform post-processing on day-ahead NWP forecasts from a single run. For example in [41], at 12:00 UTC each day, forecasts for the next 24 h are issued, but these forecasts would not be updated over the course of the day. As mentioned earlier, it can be assumed *a priori* that forecasts with short lead time are more accurate than those with long lead time. It is therefore of interest to compare the accuracy of post-processed forecasts based on day-ahead NWP forecasts to that based on forecasts with a higher update rate, such as 6 hourly or hourly. The present NWP models, namely, RAP and HRRR, offer such an opportunity.

Table 1 shows the three forecast-updating schemes, under which forecast post-processing can be conducted. One can see how the actual forecast horizons vary among the day-ahead, 6-hourly, and hourly cases, though the forecast time stamps are identical. Take for instance the forecast time stamp of 12:00. For this hour, the daily updated forecast issued at 00:00 corresponds to a horizon of $h = 12$; the 6-hourly updated forecast issued at 06:00 corresponds to a horizon of $h = 6$, and the hourly updated forecast issued at 11:00 has simply a 1-h-ahead horizon. Clearly, if the general rule-of-thumb of "near things are more related than distant things" is valid, the hourly updated forecast issued at 11:00 ought to be more accurate than the other two, which are issued at earlier times. Nonetheless, this *a priori* assumption—i.e., post-processing hourly updated NWP forecasts is more advantageous than post-processing those less frequently updated forecasts—must be verified, which constitutes the fourth aim of the article.

4. Data description

All forecasting works ought to involve at least two datasets, one measured and the other forecast, such that the accuracy of the forecast dataset can be quantified by comparing it to the measured one. This section provides a brief description of the ground-based global horizontal irradiance (GHI) measurements, as well as the GHI forecasts, that are used in this work. In order to ensure full reproducibility, both the datasets themselves and the R code are offered as supplementary materials, which can be found in Appendix A, alongside with some instructions to use them.

4.1. Ground-based data

In the empirical part of the article, data from the Surface Radiation Budget Network (SURFRAD) is considered as the ground-truth for gauging the performance of (post-processed) NWP forecasts. It is believed that solar forecasters are, by now, familiar with SURFRAD, for it can be easily said to be the Rolls Royce of irradiance monitoring.

Table 1
The correspondence between the forecast time stamp and forecast issuing time, over the course of the day, for the cases of (1) daily updated, (2) 6-hourly updated, and (3) hourly updated NWP forecasting systems. The respective forecast horizons are indicated in the parentheses. The day-ahead forecasts are assumed to be issued daily at 00 UTC, whereas the 6-hourly forecasts are assumed to be issued four times daily at 00, 06, 12, and 18 UTC (as is the case of NAM).

fcst. time stamp	fcst. issuing time (fcst. horizon)			fcst. time stamp	fcst. issuing time (fcst. horizon)		
	Day-ahead	6-hourly	Hourly		Day-ahead	6-hourly	Hourly
01:00	00:00 ($h = 1$)	00:00 ($h = 1$)	00:00 ($h = 1$)	13:00	00:00 ($h = 13$)	12:00 ($h = 1$)	12:00 ($h = 1$)
02:00	00:00 ($h = 2$)	00:00 ($h = 2$)	01:00 ($h = 1$)	14:00	00:00 ($h = 14$)	12:00 ($h = 2$)	13:00 ($h = 1$)
03:00	00:00 ($h = 3$)	00:00 ($h = 3$)	02:00 ($h = 1$)	15:00	00:00 ($h = 15$)	12:00 ($h = 3$)	14:00 ($h = 1$)
04:00	00:00 ($h = 4$)	00:00 ($h = 4$)	03:00 ($h = 1$)	16:00	00:00 ($h = 16$)	12:00 ($h = 4$)	15:00 ($h = 1$)
05:00	00:00 ($h = 5$)	00:00 ($h = 5$)	04:00 ($h = 1$)	17:00	00:00 ($h = 17$)	12:00 ($h = 5$)	16:00 ($h = 1$)
06:00	00:00 ($h = 6$)	00:00 ($h = 6$)	05:00 ($h = 1$)	18:00	00:00 ($h = 18$)	12:00 ($h = 6$)	17:00 ($h = 1$)
07:00	00:00 ($h = 7$)	06:00 ($h = 1$)	06:00 ($h = 1$)	19:00	00:00 ($h = 19$)	18:00 ($h = 1$)	18:00 ($h = 1$)
08:00	00:00 ($h = 8$)	06:00 ($h = 2$)	07:00 ($h = 1$)	20:00	00:00 ($h = 20$)	18:00 ($h = 2$)	19:00 ($h = 1$)
09:00	00:00 ($h = 9$)	06:00 ($h = 3$)	08:00 ($h = 1$)	21:00	00:00 ($h = 21$)	18:00 ($h = 3$)	20:00 ($h = 1$)
10:00	00:00 ($h = 10$)	06:00 ($h = 4$)	09:00 ($h = 1$)	22:00	00:00 ($h = 22$)	18:00 ($h = 4$)	21:00 ($h = 1$)
11:00	00:00 ($h = 11$)	06:00 ($h = 5$)	10:00 ($h = 1$)	23:00	00:00 ($h = 23$)	18:00 ($h = 5$)	22:00 ($h = 1$)
12:00	00:00 ($h = 12$)	06:00 ($h = 6$)	11:00 ($h = 1$)	00:00 (next day)	00:00 ($h = 24$)	18:00 ($h = 6$)	23:00 ($h = 1$)

Table 2
Metadata of the SURFRAD network. KGC stands for Köppen–Geiger climate classification.

abbrv.	Station	Latitude	Longitude	Elevation	Time zone	KGC	After-QC avail.
BON	Bondville, IL	40.05192°N	88.37309°W	230 m	Central (GMT−6)	Cfa	98.47
DRA	Desert Rock, NV	36.62373°N	116.01947°W	1007 m	Pacific (GMT−8)	BWk	99.05
FPK	Fort Peck, MT	48.30783°N	105.10170°W	634 m	Mountain (GMT−7)	BSk	93.03
GWN	Goodwin Creek, MS	34.2547°N	89.8729°W	98 m	Central (GMT−6)	Cfa	97.60
PSU	Penn. State Univ., PA	40.72012°N	77.93085°W	376 m	Eastern (GMT−5)	Cfb	97.50
SXF	Sioux Falls, SD	43.73403°N	96.62328°W	473 m	Central (GMT−6)	Dfa	96.92
TBL	Table Mountain, CO	40.12498°N	105.23680°W	1689 m	Mountain (GMT−7)	BSk	98.33

In comparison to data from the others stations of the Baseline Surface Radiation Network (BSRN), SURFRAD data is of a higher after-quality-control (QC) availability. Moreover, SURFRAD data is made publicly available at a higher frequency than BSRN, which usually suffers from an approximately one-year delay.

There are a total of seven stations in SURFRAD, covering five major climate classes in the contiguous United States. The metadata of SURFRAD is provided in Table 2. Since both RAP and HRRR forecasts come from the year 2020—see below for more details—the corresponding measurements from the same year is retrieved. A basic QC step is taken in order to ensure a baseline quality of the ground truth, that is, the extreme-rare limits test, which filters out those data points that can be deemed highly unlikely to occur in nature. This test is implemented in the `SolarData` package of R [43,44], which also offers functions to download and read the raw SURFRAD data files into the R environment.

Once QC is performed on the raw 1-min SURFRAD data, they are aggregated to 1-h resolution. It should be noted, however, that the aggregation uses a "center" scheme, as opposed to the usual "ceiling" scheme. This is owing to the fact that NWP output represents the values of variables *at* the time stamp, but not the aggregate or average over the previous hour. For example, to correspond to a GHI forecast stamped with 12:00, one must aggregate 1-min SURFRAD data from 11:31 to 12:30 (i.e., center to the time stamp of interest), but not from 11:01 to 12:00 (ceiling), nor from 12:00 to 12:59 (floor). The reader is referred to Yang [45] for additional information related to aggregation schemes and their implications on verification.

4.2. RAP and HRRR data

Forecasts from RAP and HRRR models are used in this work. Both RAP and HRRR are operational models, of which the forecasts are disseminated, as soon as they become available, through the NOAA Operational Model Archive and Distribution System (NOMADS).[5] However, due to storage concerns, forecasts for the operating day are held on the server for a few days only, depending on the model—e.g., NOMADS holds NAM forecasts for the past seven days, whereas it holds RAP and HRRR only for a day. In this regard, one must collect data on a daily basis through an automated downloader.

RAP and HRRR provide a wealth of weather variables which can be selected either manually or by programming means. The variable that corresponds to GHI in these forecasting systems is named as "surface downward short-wave radiation flux" or `dswrfsfc`; however, one should note that the same variable takes the name `DSWRF` in the `GRIB2` files—a concise data format commonly used in meteorology to store weather forecasts. The output of RAP and HRRR can be accessed by the `rNOMADS` package in R [46]. For automated data downloading, an R script is written, which is simply a wrapper to the functions in the `rNOMADS` package, and hosted it on Amazon Web Services (AWS); interested readers may contact the corresponding author should they wish to obtain the script for their own data collection. Once the `GRIB2` files of each day are fetched from the NOMADS server, the R script extracts forecasts at the SURFRAD locations, and stores them as `txt` files for easy manipulation.

One important concern in regard to NWP models is their versioning. Given the complexity of NWP models, even a small change in parameterization schemes may result in very distinct characteristics of the model output. For this reason, when conducting NWP forecast verification, it is preferred to restrict the exercise to those forecasts from a "frozen model", that is, a specific version. Both RAP and HRRR, as mentioned earlier, had gone through several upgrades over the past years, with the most recent one took place on December 2, 2020. Therefore, before that date, RAP was in version 4 (RAPv4) whereas HRRR was in version 3 (HRRRv3); on and after that, they became RAPv5 and HRRRv4, respectively. In this article, the period of interest is January 1 to December 31, 2020, which covers one full year, but spans two versions. Nonetheless, the effect on verification due to the relatively few forecasts from RAPv5 and HRRRv4, in comparison to the number of forecasts from RAPv4 and HRRRv3, may be considered as uncritical to the present goals.

Last but not least, both the ground-based measurements and NWP forecasts contain gaps. Whereas the gaps in the former are due to instrumentation error and communication failure, those in the latter

[5] https://nomads.ncep.noaa.gov/.

are caused by occasional NOMADS server down time, during which forecasts are not disseminated. Since these gaps will affect the training of time series models—most software implementations of time series models require complete data—they are filled, whenever necessary, with McClear clear-sky irradiance [47,48], which is the recommended clear-sky irradiance for solar forecasting works [49]. Notwithstanding, time stamps with either missing forecast or missing observation are excluded during verification.

5. Verification

This section is concerned with verifying forecasts from the two NWP models in their raw forms. The first half of this section deals with absolute verification, in which the performance RAP and HRRR is assessed individually, through the IRA metric as described in Eq. (2). The second half of this section uses relative forecast verification, where the performance of RAP and HRRR are compared to that of two time series forecasting methods, namely, the autoregressive integrated moving average (ARIMA) and exponential smoothing (ETS) families of models, via the skill score.

5.1. Internal relative accuracy

When [30] designed their measure of predictability, it was intended to be general, such that one can adopt any loss function, which would grant the measure some flexibility. Following that, the IRA can also be used in the same fashion. Since the root mean square error (RMSE) is by far the most popular metric for deterministic solar forecast verification [4], it is used here. For each of the seven SURFRAD stations and for each forecast horizon, the RMSEs of RAP and HRRR are plotted in the top row of Fig. 1, whereas the IRAs are plotted in the bottom row—they in fact depict the same information.

One is able to conclude immediately from the RMSE plots that the rate at which the error increases with horizon is faster at some stations but not others. For instance, at the DRA station, which has the lowest RMSE among all stations, RMSE is near constant for all horizons. This may point at the fact that the solar irradiance at DRA is "trivial" to forecast, since clear-sky conditions predominate in Desert Rock, Nevada. It further implies that the room for improvement on forecast skill at such locations may be limited, for one may simply forecast the clear-sky irradiance and attain a satisfactory accuracy—even if a more sophisticated method is used, its complexity would not be justified, given the marginal improvement in accuracy it brings to the table. In contrast, the RMSEs for RAP forecasts at BON and GWN increase about 20 W/m^2 from the shortest to the longest horizon, indicating a loss of skill for long-range forecasting. In this it suggests that it is indeed worth investigating more efforts on finding out why the accuracy drops, which would, in turn and in hope, lead to better forecasting techniques.

The word "predictability" has not been used to explain any observation made from Fig. 1. This is because the scientific theory on predictability with which this section is to be concerned is, in the main, a notion without consensus, although various interpretations of it have been proposed in the literature [32]. It is customary to regard weather conditions which can lead to lower RMSE (or any other negatively oriented metric) as predictable, and the converse as unpredictable. This straightforward definition is problematic in several respects. Take first its most important aspect, a high RMSE may not be due to the predictability itself, instead it may be caused by a bad forecasting model. For those who advocate using the same forecasting system to gauge predictability, it is easy to counter-argue that the system would, inevitably, favor some forecasting situations over others. Therefore, predictability must be measured internal to a particular forecasting situation, with a decent forecasting system, as is how [30] designed their measure to be.

Secondly, predictability should be associated with the potential amount of improvement that one is able to attain on top of that achieved by a naïve reference method [32]. When one thinks of a person being skillful, it is necessary to look into those difficulties which no ordinary man is unable to address. Similarly, a skilled forecaster should be able to improve upon the accuracy reached by the no-so-skilled forecasters. Since irradiance under clear skies does not require skill to forecast, and no forecaster is able to improve forecast accuracy substantially in this situation, there is no predictability to it. This argument can be extended to all those situations under which even the best forecasters are incapable to achieve more than a good educated guess, such as forecasting a white noise series.

Combining the two aspects, it appears that IRA$(1, h_{\max})$, computed based on RAP or HRRR forecasts, can be viewed as a rough measure of predictability. In that, it compares the errors achieved by the same, decent forecasting system at short and long horizons. For the irradiance at the DRA station, the predictability is low, since its IRA$(1, h_{\max}) \approx 0$. On the other hand, at the BON, GWN, and SXF stations, predictability is high, for the differences between the RMSEs at $h = 1$ and $h = 18$ or 21 are substantial. Also worth-mentioning is that IRA is found dropping as horizon increases, which echoes the fact that the forecast difficulty generally increases with forecast horizon. The result of the IRA analysis presented in this section is highly consistent with the predictability reported by Yang et al. [32], with data from the same stations but another forecasting method.

5.2. NWP forecasts versus time series forecasts

The main aim of having hourly updated NWP forecasts is to facilitate short-term decision-making, which requires the most recent, and hopefully more accurate, information in regard to the quantity being forecast. NWP forecasts are necessarily more advantageous than time series or machine-learning-based forecasts for one reason: they do not require on-site irradiance measurements for forecast generation. That said, if the accuracy of such NWP forecasts is inferior to those forecasts generated by alternative means, one must either post-process the NWP forecasts or switch to the alternative forecasts for decision-making.

Two classic time series forecasting methods—ARIMA and ETS—are implemented, and both are able to handle seasonal time series. In order to generate time series forecasts in the same fashion as the RAP and HRRR forecasts, hourly rolling, multiple-step-ahead forecasting is conducted. The procedure consists of two loops:

1 Outer loop: For each forecasting day, the process order of an ARIMA (or ETS) model is estimated using 14 previous days' information up to 23:00 on the day prior to the forecasting day, e.g., data from January 1–14 is used to determine the model for January 15, and January 2–15 is used to determine the model for January 16, so on and so forth.
2 Inner loop:
   (a) For each hour of the forecasting day, the model is refitted with the latest information, but the process order is not re-estimated.[6]
   (b) 21-step-ahead ARIMA and ETS forecasts are generated at each forecasting-issuing hour.

The whole process is iterated over the year. For rolling forecasts, the reader is referred to https://robjhyndman.com/hyndsight/rolling-forecasts/, as well as the supplementary materials in Appendix A. In this article, the `auto.arima` and `ets` functions in the `forecast` package in R [50,51] are used for daily process-order estimation, and `Arima` and `ets` functions are used for hourly refitting and forecasting.

[6] To construct a time series model, both the process order (i.e., the function form) and the exact coefficients are needed. Here in the inner loop, only coefficients changes with each iteration, but the process order remains unchanged throughout each inner loop.

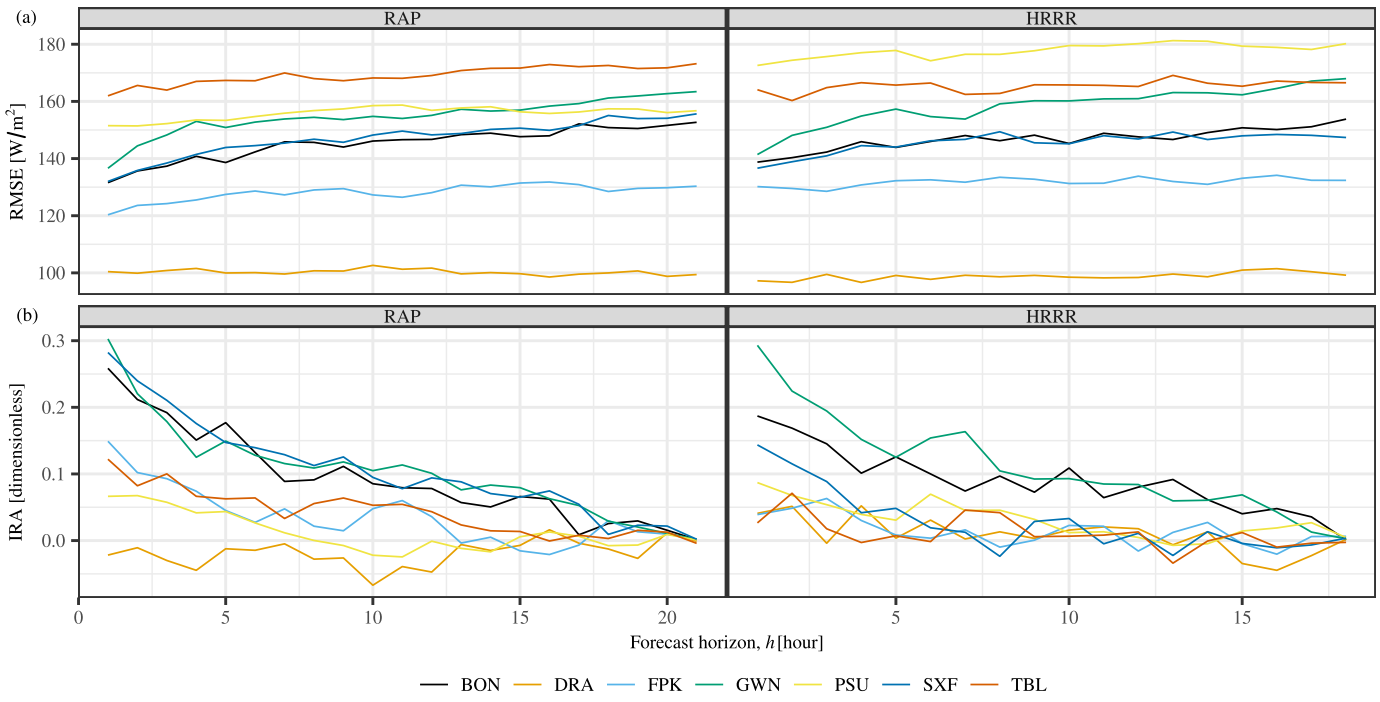


Fig. 1. (a) RMSE, in W/m^2, of h-step-ahead RAP and HRRR forecasts at seven SURFRAD stations, over the year 2020. (b) Internal relative accuracy of RAP and HRRR.

Besides ARIMA and ETS, a naïve reference method, namely, the optimal convex combination of climatology and persistence, is also implemented in a like fashion. Both the climatology forecast (i.e., the mean clear-sky index), μ, and the lag-h autocorrelation, γ_h, are derived from data from the 14 previous days, and the reference h-step-ahead forecast for time t is simply $\gamma_h x_{t-h} + (1 - \gamma_h)\mu$. Furthermore, clear-sky indexes of the nighttime hours are assigned with the annual mean value. Though this treatment can be regarded as artificial, since the entire literature on univariate time series solar forecasting is faced with the same challenge and does not offer any better alternative, one has no choice but to assume a "no change" transition from late afternoon to early morning is admissible. The situation can be ameliorated when nighttime cloud transients are included as an exogenous variable, as how it would be used in a multivariate time series forecasting setting [see 52,53, for ARIMA and ETS modeling with nighttime cloud index]. That said, such an approach can no longer be viewed as naïve, and therefore is irrelevant to our current discussion.

Fig. 2 shows the RMSE skill scores of ARIMA, ETS, RAP, and HRRR for horizon ranges from 1- to 18-h ahead, evaluated based on all forecasts that correspond to a zenith angle $\theta < 85°$. It is "surprising" to see that the classic time series modeling does not lead to positive skill score at all stations, contradicting what so far has been reported in the literature—indeed, there have been numerous cases studies showing time series modeling, e.g., autoregressive models [54], ETS models [55], or k-nearest neighbor [56], is able to achieve a positive skill score. The obvious explanation for the contradiction between current results and the ones from the literature is that the optimal convex combination of climatology and persistence is far more accurate than climatology or persistence, on which the skill scores reported in the literature are based. As argued by Murphy [57], when multiple naïve reference methods are present, the one which returns the lowest error should be used as the standard of reference, the entire literature on solar forecasting might have been exaggerating skill score due to the lack of awareness of this more advanced naïve reference method.

Another interesting observation is that the skill score of ARIMA and ETS drops at long horizons beyond 15 h. The underlying rationale is thought to be one related to climatology and its long-range optimality. The reference forecast, $\gamma_h x_{t-h} + (1 - \gamma_h)\mu$, depends on the lag-$h$ autocorrelation, γ_h. When h is large, such as $h \geq 15$, γ_h is often near zero,[7] which in turn makes the climatology component the dominating one in the convex combination. This simple fact that climatology forecast outperforms the forecasts issued by the more elaborate ARIMA and ETS models indicates a potential "white noise" nature of solar irradiance at long range, in which issuing any forecast besides the mean would lead to inferior performance, insofar as univariate time series forecasting is concerned.

Having examined the behavior of time series forecasts, NWP forecasts are inspected next. The forecast skill of the two NWP models, like that of time series methods, is unsatisfactory in general. Particularly curious is the poor performance at short horizons, which suggests possibly an unsatisfactory analysis that is substantially biased. (This is further investigated in the next section.) At horizons beyond 5 h, NWP forecasts seem more advantageous than the time series methods at three out of seven stations, namely, BON, GWN, and SXF. Coincidentally or not, these three stations are also found to possess the highest internal relative accuracy in the earlier section. What is not clear at this stage is whether this good performance is due to the inherent high predictability at the three stations, or is due to the fact that the modeling and setup of RAP and HRRR favor the climate and weather regimes at these three stations, or perhaps, is the effect of two causes, simultaneously.

6. Post-processing

Post-processing of NWP forecasts is a standard practice. Indeed, post-processing has been studied extensively [12], it is not possible, at this date, to say anything very new about it. Nevertheless, it is necessary to describe it, each time, if we are to be in a position later to consider what constitutes the most beneficial post-processing technique for the forecasts at hand.

In selecting post-processing methods there are two main stages: the first consists in identifying the category of forecast conversion; the

[7] This claim is supported empirically by plotting out the autocorrelation function of clear-sky index time series, which one can verify using the data provided in the supplementary materials.

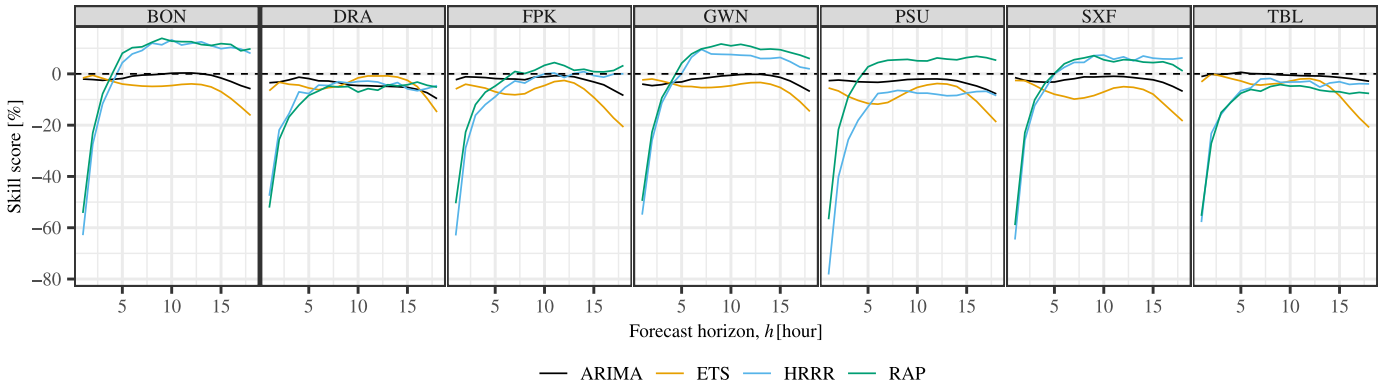


Fig. 2. Skill scores, in percent, of ARIMA, ETS, RAP, and HRRR, at seven SURFRAD stations, over the year 2020, for horizons $h = 1, \ldots, 18$. The optimal convex combination of climatology and persistence is used as the standard of reference in skill score computation.

second in choosing a "thinking tool" within the selected category. As noted by Yang and van der Meer [12], all forecast post-processing must belong to one of the four categories, namely, deterministic to deterministic (D2D), deterministic to probabilistic (D2P), probabilistic to deterministic (P2D), or probabilistic to probabilistic (P2P). For each category, there are two to three thinking tools that can be deemed as overarching, so far as their ability to summarize the specific post-processing methods is considered. In this article, it is the D2D post-processing that is of interest, and there are three thinking tools in that category: (1) performing regression, (2) performing filtering, (3) changing the forecast resolution [12].

Regression aims at constructing a regressive relationship between the bias of NWP forecasts and various explanatory variables. In that, most statistical and machine-learning-based post-processing methods, such as one using a neural network, support vector machine, regression tree, or model output statistics, are essentially performing regression. On the other hand, filtering, though similar to regression in form, is profoundly different from it in substance. It is a sequential procedure which aims at stabilizing forecasts over several forecast horizons. The initial few corrections made to the raw forecasts may be bad, but the corrective ability tends to improve over time. The reader is referred to Yang and van der Meer [12] for the distinction between regression and filtering in post-processing. In what follows, this article shall only be concerned with Kalman filtering, while noting there are alternatives that can also be deemed suitable.

6.1. Kalman filtering

It would not be difficult to find definitions of Kalman filter from textbooks. One example is this: Kalman filter is an recursive filter that estimates the state of a dynamical system from a series of noisy measurements. This kind of definitions and the subsequent mathematics are, however, abstract and general. Though useful, they must not be regarded as instructional, for the ones who are not already familiar with the topic may easily get carried away, particularly because, at present, countless versions of text on Kalman filtering seldom share any consistent terminology or symbols. To that end, this section introduces Kalman filtering from a solar forecasting perspective, in a restricted but comprehensible way.

6.1.1. General modeling

The mathematical detail about the Kalman filtering procedure is first provided. For clarity, the following conventions are adopted: (1) matrices and vectors are written in bold; (2) column vectors are written as lower-case letters; and (3) when a vector is accompanied by a superscript "$\top$", denoting the transpose, it is a row vector.

The Kalman filtering procedure assumes the vector of variables of interest, $\boldsymbol{y}_t \in \mathbb{R}^{p\times 1}$, can be expressed as a linear model of a hidden (i.e., unknown) *state vector*, $\boldsymbol{\alpha}_t \in \mathbb{R}^{m\times 1}$, and a known *measurement matrix*, $\boldsymbol{Z}_t \in \mathbb{R}^{p\times m}$, representing the noiseless connection between the state vector and the variables themselves, that is,

$$\boldsymbol{y}_t = \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \tag{6}$$

where $\boldsymbol{\varepsilon}_t \in \mathbb{R}^{p\times 1}$ is an error term, which is often taken to be a white-noise process with covariance matrix $\boldsymbol{H}_t \in \mathbb{R}^{p\times p}$. Eq. (6) is called the *measurement equation*, describing how the observed values of the variables of interest relate to the state vector.

Similar to the variable of interest, the state vector $\boldsymbol{\alpha}_t$ can also be modeled linearly, through a *state equation*, describing how it is evolved through time:

$$\boldsymbol{\alpha}_t = \boldsymbol{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \tag{7}$$

which suggests $\boldsymbol{\alpha}_t$ can be evolved from $\boldsymbol{\alpha}_{t-1} \in \mathbb{R}^{m\times 1}$ through a *state transition matrix* $\boldsymbol{T}_t \in \mathbb{R}^{m\times m}$, and $\boldsymbol{\eta}_t \in \mathbb{R}^{m\times 1}$ is a vector white-noise process with covariance matrix $\boldsymbol{Q}_t \in \mathbb{R}^{m\times m}$. At this stage, it should be made clear that the goal of Kalman filtering is to estimate $\boldsymbol{\alpha}_t$ using the information up to time $t-1$, in that, the estimated state vector should be denoted as $\hat{\boldsymbol{\alpha}}_{t|t-1}$. Once $\hat{\boldsymbol{\alpha}}_{t|t-1}$ is known, $\hat{\boldsymbol{y}}_{t|t-1} = \boldsymbol{Z}_t\hat{\boldsymbol{\alpha}}_{t|t-1}$ becomes available, which contains the predicted values of variables of interest—recall that the measurement matrix $\boldsymbol{Z}_t$ is known.

The recursive procedure of Kalman filtering proceeds as follows. Firstly, initial values for $\hat{\boldsymbol{\alpha}}_{0|0}$ and $\boldsymbol{P}_{0|0}$ are assigned, and the prediction step is given as:

$$\hat{\boldsymbol{\alpha}}_{t|t-1} = \boldsymbol{T}_t\hat{\boldsymbol{\alpha}}_{t-1|t-1}, \tag{8}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{T}_t^\top + \boldsymbol{Q}_t, \tag{9}$$

where $\boldsymbol{T}_t$ and $\boldsymbol{Q}_t$ can be either assumed to be known or estimated through statistical procedures. In other words, with $\hat{\boldsymbol{\alpha}}_{0|0}$ and $\boldsymbol{P}_{0|0}$, $\hat{\boldsymbol{\alpha}}_{1|0}$ and $\boldsymbol{P}_{1|0}$ are obtained. Subsequently, when the measurements $\boldsymbol{y}_t$ becomes available, the system can be updated through the *update equations*:

$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t^\top\left(\boldsymbol{Z}_t\boldsymbol{P}_{t|t-1}\boldsymbol{Z}_t^\top + \boldsymbol{H}_t\right)^{-1}, \tag{10}$$

$$\hat{\boldsymbol{\alpha}}_{t|t} = \hat{\boldsymbol{\alpha}}_{t|t-1} + \boldsymbol{K}_t\left(\boldsymbol{y}_t - \boldsymbol{Z}_t\hat{\boldsymbol{\alpha}}_{t|t-1}\right), \tag{11}$$

$$\boldsymbol{P}_{t|t} = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{Z}_t)\boldsymbol{P}_{t|t-1}, \tag{12}$$

where $\boldsymbol{K}_t \in \mathbb{R}^{m\times p}$ is called the *Kalman gain*, $\boldsymbol{I} \in \mathbb{R}^{m\times m}$ is an identity matrix, and $\boldsymbol{H}_t$ is similar to $\boldsymbol{Q}_t$, for which the value can be either assumed or estimated. That is, with $\hat{\boldsymbol{\alpha}}_{1|0}$, $\boldsymbol{P}_{1|0}$ and $\boldsymbol{y}_1$, $\hat{\boldsymbol{\alpha}}_{1|1}$ and $\boldsymbol{P}_{1|1}$ are obtained. The process then goes back to Eqs. (8) and (9), where $\hat{\boldsymbol{\alpha}}_{2|1}$ and $\boldsymbol{P}_{2|1}$ are obtained from $\hat{\boldsymbol{\alpha}}_{1|1}$ and $\boldsymbol{P}_{1|1}$, so on and so forth. This concludes the general Kalman filtering procedure.

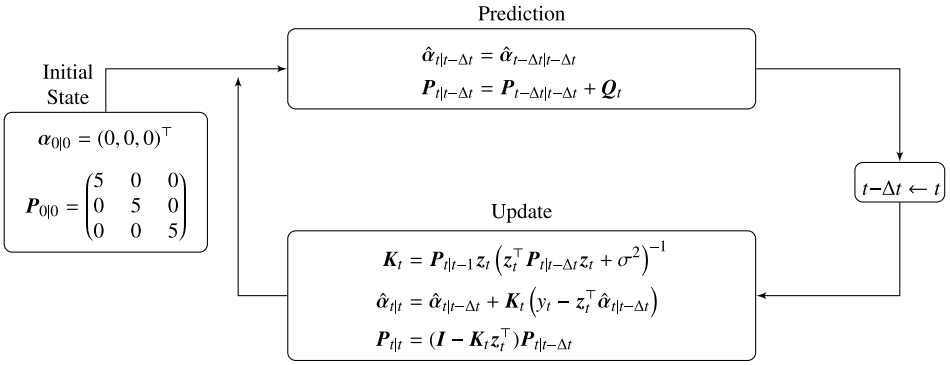


Fig. 3. The recursive procedure of Kalman filtering used in this work. A future time to be filtered is stamped t. The "prediction" stage makes estimations on this future state ($\hat{\boldsymbol{\alpha}}_{t|t-\Delta t}$) and future error covariance ($\boldsymbol{P}_{t|t-\Delta t}$), based on the current information. When time moves forward—the previous-round "future" time (or t) becomes the "current" time (or $t-\Delta t$, which lags the "new future" time by Δt)—the "update" stage updates the Kalman gain ($\boldsymbol{K}_t$), states ($\hat{\boldsymbol{\alpha}}_{t|t}$) and error covariance ($\boldsymbol{P}_{t|t}$).

6.1.2. Kalman filtering as an NWP post-processing tool

Take that one wishes to apply Kalman filtering on the time series of bias of NWP forecasts. As per the usual convention of solar forecasting, as mentioned earlier, it is more amenable to operate on clear-sky index, namely, let

$$y_t = \frac{\text{bias}_t}{c_t} = \frac{f_t - x_t}{c_t}, \tag{13}$$

where f_t and x_t are forecast and observation, respectively, and c_t is the clear-sky expectation. Additionally, when Kalman filtering is applied to NWP forecasts, a series of simplifications can be made, without dwarfing its overall ability to act as a post-processing tool [58]. To that end, this article combines the approaches that are depicted in [41,42,58], and considers a simplified version of Kalman filter, see Fig. 3.

Since the variable of interest is now a scalar quantity, the measurement equation is simplified to:

$$y_t = \boldsymbol{z}_t^\top \boldsymbol{\alpha}_t + \varepsilon_t, \tag{14}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ is a homoscedastic Gaussian noise. Furthermore, given the fact that y_t is the bias of clear-sky index forecast, it is reasonable to assume $\sigma^2 = 0.1$. The state equation does not consider any transition other than a simple innovation, in that, it is reduced to:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-\Delta t} + \boldsymbol{\eta}_t. \tag{15}$$

It should be noted that instead of annotating the previous-step state vector by "$t|t-1$", the subscript is replaced with "$t|t-\Delta t$" to handle multiple-step-ahead forecasts. For instance, when 2-h-ahead forecasts are post-processed, $\Delta t = 2$ h. The predictive equations are therefore:

$$\hat{\boldsymbol{\alpha}}_{t|t-\Delta t} = \hat{\boldsymbol{\alpha}}_{t-\Delta t|t-\Delta t}, \tag{16}$$

$$\boldsymbol{P}_{t|t-\Delta t} = \boldsymbol{P}_{t-\Delta t|t-\Delta t} + \boldsymbol{Q}_t, \tag{17}$$

and the update equations become:

$$\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{z}_t \left(\boldsymbol{z}_t^\top \boldsymbol{P}_{t|t-\Delta t}\boldsymbol{z}_t + \sigma^2\right)^{-1}, \tag{18}$$

$$\hat{\boldsymbol{\alpha}}_{t|t} = \hat{\boldsymbol{\alpha}}_{t|t-\Delta t} + \boldsymbol{K}_t\left(y_t - \boldsymbol{z}_t^\top \hat{\boldsymbol{\alpha}}_{t|t-\Delta t}\right), \tag{19}$$

$$\boldsymbol{P}_{t|t} = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{z}_t^\top)\boldsymbol{P}_{t|t-\Delta t}. \tag{20}$$

Following Diagne et al. [42], the measurement vector $\boldsymbol{z}_t$ is three-dimensional, that is,

$$\boldsymbol{z}_t = \left(1, \hat{\kappa}_t, \cos\theta_t\right)^\top, \tag{21}$$

where $\hat{\kappa}_t \equiv f_t/c_t$ is the forecast clear-sky index at time t, and θ_t is the zenith angle at time t.[8] The initial values are set as:

$$\boldsymbol{\alpha}_{0|0} = (0,0,0)^\top, \tag{22}$$

$$\boldsymbol{P}_{0|0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{23}$$

Finally, $\boldsymbol{Q}_t$ takes a constant value of

$$\boldsymbol{Q}_t = \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.05 & 0 \\ 0 & 0 & 0.05 \end{pmatrix}. \tag{24}$$

6.1.3. Filtering for multiple-step-ahead forecasts

The Kalman filtering procedure is sequential, in that, prediction equations and update equations are used iteratively. This characteristic leads to complications when the filtering procedure is applied to multiple-step-ahead forecasts. For instance, if the Δt in Eqs. (15) is 2 h, time will proceed forward two 1-h steps in each iteration, and therefore leaving "holes" in the filtered time series. This issue, unfortunately, has not been fully discussed in the solar forecasting literature. Based on what has been presented in the literature, one has reasons to suspect that so-called "multiple-step-ahead Kalman filtering" used in those works was actually 1-step-ahead filtering applied on multiple-step-ahead forecasts issued at earlier times—this ambiguity as to multiple-step-ahead filtering has also aroused interest of many other researchers (e.g., pers. comm. Pierre Pinson).

To circumvent the unnecessary hustle of investigating into how past results on multiple-step-ahead Kalman filtering were arrived, a workaround can be easily thought of. Given the fact that the filter will step forward h units of time in each iteration, the simplest way is to construct h filters, so that they will jointly cover all forecasts. To give perspective, consider a filtering task for a series of 2-h-ahead forecasts that are updated hourly—at 00:00, forecast for 02:00 is issued, and at 01:00, forecast for 03:00 is issued, so on and so forth—such as the case of RAP and HRRR. One can construct a Kalman filter using observations and forecasts made for $t = 1, 3, 5, 7, \ldots$, and another using observations and forecasts made for $t = 2, 4, 6, 8, \ldots$, such that both filters have $\Delta t = 2$. Following this strategy, for h-step-ahead forecasts filtering, h Kalman filters can be used. This has also been the strategy for multiple-step-ahead Kalman filtering used in Yang et al. [20].

[8] The original choice of state vector, as appeared in Diagne et al. [42], uses $f_t/1000$ in the place of $\hat{\kappa}_t$, which is illogical, because dividing NWP forecasts by 1000 W/m^2 does not remove the diurnal cycle in the forecasts.

Table 3
MBE, in W/m^2 (%), of raw and Kalman-filtered RAP and HRRR forecasts, for five horizons, $h = 1, 2, 3, 6, 18$ h, at seven SURFRAD stations. ARIMA and ETS forecasts results are tabulated for reference. Row-wise best results, i.e., lowest absolute MBE, are in bold.

stn.	h	raw NWP		Time series methods			Kalman-filtered NWP	
		RAP	HRRR	ARIMA	ETS	REF	RAP + KF	HRRR + KF
BON	1 h	56.28 (15.29%)	68.90 (18.72%)	2.97 (0.81%)	3.05 (0.83%)	**0.08 (0.02%)**	0.88 (0.24%)	0.98 (0.27%)
BON	2 h	59.26 (16.10%)	69.61 (18.91%)	4.43 (1.20%)	3.60 (0.98%)	0.70 (0.19%)	**0.05 (0.01%)**	0.47 (0.13%)
BON	3 h	59.04 (16.04%)	70.65 (19.19%)	5.16 (1.40%)	3.79 (1.03%)	1.73 (0.47%)	−0.87 (−0.24%)	**−0.64 (−0.17%)**
BON	6 h	60.95 (16.55%)	71.08 (19.31%)	4.88 (1.32%)	2.68 (0.73%)	**1.23 (0.33%)**	−2.64 (−0.72%)	−2.74 (−0.74%)
BON	18 h	56.77 (15.37%)	72.55 (19.65%)	2.86 (0.77%)	−10.27 (−2.78%)	**−0.56 (−0.15%)**	−2.44 (−0.66%)	−2.62 (−0.71%)
DRA	1 h	46.53 (8.89%)	44.34 (8.47%)	−1.61 (−0.31%)	2.10 (0.40%)	−1.11 (−0.21%)	1.60 (0.30%)	**1.03 (0.20%)**
DRA	2 h	46.53 (8.89%)	44.15 (8.44%)	−2.98 (−0.57%)	2.05 (0.39%)	−1.61 (−0.31%)	1.77 (0.34%)	**1.33 (0.25%)**
DRA	3 h	46.86 (8.96%)	43.90 (8.39%)	−3.67 (−0.70%)	1.87 (0.36%)	−2.30 (−0.44%)	0.82 (0.16%)	**0.39 (0.07%)**
DRA	6 h	47.37 (9.05%)	43.52 (8.32%)	−3.49 (−0.67%)	1.58 (0.30%)	−1.87 (−0.36%)	−0.68 (−0.13%)	**−0.46 (−0.09%)**
DRA	18 h	46.66 (8.91%)	43.93 (8.39%)	−5.50 (−1.05%)	−0.88 (−0.17%)	−2.12 (−0.40%)	**−0.04 (−0.01%)**	−0.37 (−0.07%)
FPK	1 h	53.04 (14.21%)	67.21 (18.01%)	2.73 (0.73%)	2.46 (0.66%)	6.00 (1.61%)	**1.50 (0.40%)**	1.60 (0.43%)
FPK	2 h	57.18 (15.33%)	65.82 (17.64%)	4.05 (1.09%)	4.12 (1.10%)	9.04 (2.42%)	1.76 (0.47%)	**0.97 (0.26%)**
FPK	3 h	59.83 (16.03%)	66.60 (17.85%)	4.81 (1.29%)	5.53 (1.48%)	10.92 (2.93%)	0.78 (0.21%)	**0.36 (0.10%)**
FPK	6 h	64.42 (17.26%)	68.26 (18.29%)	6.18 (1.66%)	8.10 (2.17%)	11.51 (3.08%)	−1.47 (−0.39%)	**−1.32 (−0.35%)**
FPK	18 h	63.47 (17.01%)	66.27 (17.76%)	5.09 (1.36%)	1.14 (0.31%)	9.81 (2.63%)	**−0.40 (−0.11%)**	−1.36 (−0.36%)
GWN	1 h	61.54 (15.97%)	65.76 (17.06%)	−0.91 (−0.24%)	1.76 (0.46%)	−0.93 (−0.24%)	0.90 (0.23%)	**0.31 (0.08%)**
GWN	2 h	58.55 (15.19%)	69.48 (18.03%)	−1.64 (−0.42%)	2.22 (0.58%)	−1.02 (−0.27%)	**−0.01 (0.00%)**	−0.06 (−0.02%)
GWN	3 h	59.31 (15.39%)	69.05 (17.91%)	−2.32 (−0.60%)	1.96 (0.51%)	−0.82 (−0.21%)	**−0.40 (−0.10%)**	−0.72 (−0.19%)
GWN	6 h	65.23 (16.93%)	69.93 (18.15%)	−3.17 (−0.82%)	−0.90 (−0.23%)	**−0.37 (−0.10%)**	−2.76 (−0.72%)	−2.68 (−0.70%)
GWN	18 h	67.96 (17.59%)	79.76 (20.65%)	−2.56 (−0.66%)	−8.53 (−2.21%)	**−0.46 (−0.12%)**	−3.24 (−0.84%)	−4.27 (−1.10%)
PSU	1 h	77.63 (22.71%)	102.98 (30.12%)	1.55 (0.45%)	2.92 (0.85%)	2.04 (0.60%)	**0.22 (0.06%)**	−0.75 (−0.22%)
PSU	2 h	78.02 (22.82%)	105.55 (30.87%)	2.25 (0.66%)	4.25 (1.24%)	2.85 (0.83%)	**−0.78 (−0.23%)**	−1.42 (−0.42%)
PSU	3 h	77.43 (22.65%)	106.54 (31.16%)	2.70 (0.79%)	5.13 (1.50%)	4.18 (1.22%)	**−1.34 (−0.39%)**	−2.37 (−0.69%)
PSU	6 h	79.19 (23.17%)	103.82 (30.37%)	2.74 (0.80%)	5.61 (1.64%)	5.28 (1.54%)	**−2.47 (−0.72%)**	−2.83 (−0.83%)
PSU	18 h	74.85 (21.87%)	105.52 (30.83%)	**2.28 (0.67%)**	−4.61 (−1.35%)	4.88 (1.43%)	−3.08 (−0.90%)	−4.23 (−1.24%)
SXF	1 h	60.91 (16.40%)	67.07 (18.06%)	1.16 (0.31%)	1.03 (0.28%)	2.06 (0.56%)	**−0.24 (−0.06%)**	−0.49 (−0.13%)
SXF	2 h	62.63 (16.87%)	68.69 (18.50%)	1.87 (0.50%)	**0.81 (0.22%)**	3.18 (0.86%)	−0.84 (−0.23%)	−1.09 (−0.29%)
SXF	3 h	64.68 (17.42%)	69.88 (18.82%)	2.40 (0.65%)	**0.43 (0.11%)**	3.84 (1.03%)	−1.05 (−0.28%)	−1.52 (−0.41%)
SXF	6 h	67.53 (18.18%)	72.31 (19.47%)	3.78 (1.02%)	**−1.31 (−0.35%)**	4.56 (1.23%)	−2.06 (−0.55%)	−1.94 (−0.52%)
SXF	18 h	60.51 (16.33%)	71.35 (19.25%)	2.86 (0.77%)	−3.11 (−0.84%)	5.15 (1.39%)	−3.62 (−0.98%)	**−2.55 (−0.69%)**
TBL	1 h	83.05 (19.22%)	79.05 (18.29%)	**4.35 (1.01%)**	6.01 (1.39%)	6.51 (1.51%)	7.08 (1.64%)	6.90 (1.60%)
TBL	2 h	82.76 (19.15%)	78.01 (18.05%)	**4.75 (1.10%)**	8.31 (1.92%)	7.88 (1.82%)	7.01 (1.62%)	6.89 (1.59%)
TBL	3 h	84.12 (19.47%)	81.16 (18.78%)	**3.86 (0.89%)**	9.65 (2.23%)	7.34 (1.70%)	5.58 (1.29%)	5.35 (1.24%)
TBL	6 h	83.03 (19.21%)	80.05 (18.52%)	**0.37 (0.09%)**	8.41 (1.95%)	3.30 (0.76%)	−2.31 (−0.53%)	−2.44 (−0.56%)
TBL	18 h	82.87 (19.18%)	81.38 (18.83%)	−2.48 (−0.57%)	−9.19 (−2.13%)	1.77 (0.41%)	**0.14 (0.03%)**	0.37 (0.09%)

6.2. MBE and RMSE of filtered forecasts

Based on the procedure outlined in the previous section, Kalman filters are constructed based on RAP and HRRR forecasts. Instead of overloading the paper, only results of five horizons, namely, $h = 1, 2, 3, 6, 18$ are reported, as shown in Tables 3 and 4. Nonetheless, the reader is able to view the results for other horizons by modifying the code as provided in the supplementary materials.

The MBEs of filtered RAP and HRRR forecasts are contrasted to those of raw NWP forecasts and time series forecasts in Table 3. As hypothesized, the raw NWP forecasts are severely biased at all stations for all horizons. More specifically, the bias is a positive one, indicating that situations of over-estimation are common. In contrast, the bias of time series forecasts and that of the filtered NWP forecasts are comparable, both rarely exceed 2%. On this point, the bias-removal capability of Kalman filtering can be confirmed.

Table 4 depicts the result in terms of RMSE. Overall, it is evident that performing Kalman filtering is able to lower the squared error of NWP forecasts substantially at short horizons, but not at long horizons. Furthermore, although filtered NWP forecasts are found to have attained the lowest RMSE at some stations/horizons, such good performance is not universal, as the optimal reference method is still dominating at many situations. Comparing to ARIMA and ETS, the filtered forecasts seem more advantageous, but the improvement is marginal. These observations based on the reported RMSEs echoes the earlier conclusion on the possible loss of predictability at long horizons. For horizons $h = 6$ and 18, applying Kalman filtering removes the bias but increases the variance, such that the overall RMSE remains at about the same level as that of the pre-filtering one.

6.3. Kalman filtering for NWP forecasts with different update rate

In Section 3.4, particularly, Table 1, it has been made clear that an NWP forecast time series can be constructed in various ways. It is therefore of interested to examine the accuracy of filtered forecasts that are generated by a less frequently update NWP model. For HRRR, the precise setting of Table 1 can be used, which results in three versions of NWP forecasts: hourly updated, 6-hourly updated, and daily updated. On the other hand, for RAP, since the forecasts issued at 00z only goes out to 21 h, and do not cover the next 24 h, the 03z forecasts are used. It has been empirically demonstrated that Kalman filtering is able to remove bias more generally, therefore, to save space, only the RMSE table will be displayed and discussed, whereas the reader can inspect the MBE table on his own, by running the code provided in the supplementary materials.

Table 5 shows the RMSEs of Kalman-filtered NWP forecasts with different update rates. The overall observation is this: filtering hourly updated NWP forecasts is able to lead to marginal improvement from filtering 6-hourly updated NWP forecasts, which is able to be assessed as slightly better than filtering day-ahead NWP forecasts. Considering the number of forecast–observation pairs that is involved during verification of each case—approximately 4000 pairs—the materialized improvements from one column to another in Table 5 might turn out to be statistically significant, if one conduct formal hypothesis tests. However, even if the improvements can be tested to be statistically significant, their operational impact would no doubt be limited. For that reason, hypothesis testing is omitted. Instead, it is herein concluded that, at the present level of development of dynamical weather modeling, the NWP update rate has not yet a significant impact on hour-ahead solar forecasting accuracy.

Table 4
Same as Table 3, but for RMSE in W/m^2 (%). Row-wise best results, i.e., lowest RMSE, are in bold.

stn.	*h*	raw NWP		time series methods			Kalman-filtered NWP	
		RAP	HRRR	ARIMA	ETS	REF	RAP + KF	HRRR + KF
BON	1 h	132.56 (36.01%)	140.01 (38.03%)	87.68 (23.82%)	87.31 (23.72%)	85.95 (23.35%)	84.65 (22.99%)	**83.29 (22.62%)**
BON	2 h	136.66 (37.12%)	141.49 (38.43%)	113.53 (30.84%)	111.69 (30.34%)	111.07 (30.17%)	**107.12 (29.10%)**	107.17 (29.11%)
BON	3 h	138.23 (37.54%)	143.18 (38.89%)	131.51 (35.72%)	130.49 (35.44%)	128.28 (34.84%)	**121.12 (32.90%)**	122.10 (33.16%)
BON	6 h	143.02 (38.85%)	146.93 (39.91%)	160.68 (43.64%)	166.30 (45.17%)	159.30 (43.27%)	143.72 (39.04%)	**141.97 (38.56%)**
BON	18 h	**151.59 (41.06%)**	154.54 (41.86%)	177.75 (48.14%)	195.22 (52.87%)	168.00 (45.50%)	158.09 (42.82%)	155.28 (42.06%)
DRA	1 h	101.58 (19.42%)	98.50 (18.83%)	69.07 (13.20%)	71.13 (13.60%)	**66.76 (12.76%)**	67.07 (12.82%)	67.18 (12.84%)
DRA	2 h	100.99 (19.30%)	97.98 (18.73%)	82.95 (15.85%)	83.03 (15.87%)	**80.40 (15.37%)**	81.49 (15.58%)	81.44 (15.57%)
DRA	3 h	101.94 (19.48%)	100.84 (19.27%)	89.34 (17.08%)	90.83 (17.36%)	**87.29 (16.68%)**	90.95 (17.38%)	91.18 (17.43%)
DRA	6 h	101.26 (19.35%)	99.11 (18.94%)	97.53 (18.64%)	100.82 (19.27%)	**94.95 (18.14%)**	103.22 (19.72%)	102.08 (19.51%)
DRA	18 h	101.17 (19.32%)	100.60 (19.22%)	105.49 (20.15%)	110.50 (21.11%)	**96.19 (18.37%)**	110.48 (21.10%)	110.99 (21.20%)
FPK	1 h	121.24 (32.49%)	131.35 (35.20%)	82.42 (22.09%)	85.32 (22.87%)	80.58 (21.60%)	**80.19 (21.49%)**	81.85 (21.94%)
FPK	2 h	124.50 (33.37%)	130.76 (35.04%)	102.64 (27.51%)	105.59 (28.30%)	**101.49 (27.20%)**	102.66 (27.51%)	101.98 (27.33%)
FPK	3 h	125.18 (33.55%)	129.83 (34.79%)	113.19 (30.34%)	117.14 (31.40%)	**111.79 (29.96%)**	114.03 (30.56%)	113.41 (30.39%)
FPK	6 h	129.65 (34.75%)	133.92 (35.89%)	129.64 (34.74%)	137.18 (36.76%)	**127.15 (34.08%)**	128.54 (34.45%)	130.86 (35.07%)
FPK	18 h	**129.40 (34.67%)**	133.76 (35.84%)	145.04 (38.86%)	161.51 (43.27%)	133.79 (35.85%)	136.94 (36.69%)	141.53 (37.92%)
GWN	1 h	137.93 (35.79%)	142.89 (37.07%)	95.92 (24.89%)	94.42 (24.50%)	**92.25 (23.93%)**	92.85 (24.09%)	93.41 (24.24%)
GWN	2 h	145.60 (37.78%)	149.50 (38.79%)	124.11 (32.20%)	121.04 (31.40%)	**118.67 (30.79%)**	119.48 (31.00%)	119.86 (31.10%)
GWN	3 h	149.47 (38.78%)	152.20 (39.49%)	142.51 (36.97%)	140.62 (36.48%)	136.77 (35.49%)	136.50 (35.41%)	**134.87 (34.99%)**
GWN	6 h	**153.88 (39.93%)**	155.98 (40.48%)	170.27 (44.18%)	175.03 (45.42%)	166.81 (43.28%)	157.43 (40.85%)	156.90 (40.71%)
GWN	18 h	**162.76 (42.13%)**	169.65 (43.92%)	184.81 (47.84%)	198.33 (51.34%)	173.01 (44.79%)	169.33 (43.84%)	171.80 (44.47%)
PSU	1 h	151.71 (44.37%)	172.57 (50.47%)	99.38 (29.07%)	102.14 (29.87%)	**96.82 (28.32%)**	99.05 (28.97%)	99.12 (28.99%)
PSU	2 h	151.52 (44.31%)	174.52 (51.04%)	127.48 (37.28%)	132.78 (38.83%)	**124.48 (36.41%)**	125.49 (36.70%)	126.70 (37.05%)
PSU	3 h	152.38 (44.57%)	175.98 (51.47%)	143.84 (42.07%)	152.45 (44.59%)	**140.06 (40.96%)**	140.32 (41.04%)	144.69 (42.32%)
PSU	6 h	**155.22 (45.41%)**	174.80 (51.13%)	167.72 (49.06%)	181.54 (53.11%)	162.38 (47.50%)	155.46 (45.48%)	159.48 (46.65%)
PSU	18 h	**157.94 (46.15%)**	180.92 (52.86%)	179.81 (52.54%)	198.13 (57.89%)	166.80 (48.74%)	162.20 (47.39%)	171.68 (50.16%)
SXF	1 h	132.81 (35.77%)	137.51 (37.04%)	84.75 (22.83%)	85.68 (23.08%)	**83.55 (22.50%)**	84.87 (22.86%)	85.31 (22.98%)
SXF	2 h	136.86 (36.86%)	140.13 (37.74%)	114.14 (30.74%)	114.27 (30.77%)	111.32 (29.98%)	**110.07 (29.64%)**	111.12 (29.93%)
SXF	3 h	139.54 (37.58%)	142.33 (38.33%)	130.57 (35.16%)	132.39 (35.65%)	126.73 (34.13%)	122.10 (32.88%)	**121.85 (32.82%)**
SXF	6 h	145.67 (39.23%)	147.71 (39.78%)	155.22 (41.80%)	164.99 (44.43%)	151.63 (40.83%)	**142.45 (38.36%)**	143.78 (38.72%)
SXF	18 h	156.15 (42.13%)	**148.08 (39.95%)**	168.63 (45.50%)	186.95 (50.44%)	157.92 (42.61%)	158.41 (42.74%)	153.59 (41.44%)
TBL	1 h	163.77 (37.90%)	166.34 (38.49%)	106.49 (24.64%)	108.70 (25.15%)	**105.42 (24.39%)**	106.40 (24.62%)	108.71 (25.16%)
TBL	2 h	167.53 (38.77%)	162.41 (37.58%)	132.01 (30.55%)	**131.80 (30.50%)**	131.80 (30.50%)	132.82 (30.74%)	132.46 (30.65%)
TBL	3 h	165.87 (38.38%)	166.90 (38.62%)	144.52 (33.44%)	145.38 (33.64%)	**144.17 (33.36%)**	147.86 (34.22%)	147.77 (34.19%)
TBL	6 h	169.23 (39.15%)	168.25 (38.92%)	**159.35 (36.86%)**	165.92 (38.38%)	159.53 (36.91%)	172.61 (39.93%)	172.63 (39.93%)
TBL	18 h	174.65 (40.41%)	168.61 (39.02%)	166.91 (38.62%)	196.15 (45.39%)	**162.28 (37.55%)**	184.34 (42.66%)	178.80 (41.37%)

Table 5
RMSE, in W/m^2 (%), of filtered ($\Delta t = 1$) NWP forecasts with different update rates.

stn.	RAP + KF			HRRR + KF		
	Hourly	6-hourly	Day-ahead (03z run)	Hourly	6-hourly	Day-ahead (00z run)
BON	83.93 (23.21%)	83.73 (23.15%)	84.80 (23.45%)	82.52 (22.82%)	83.94 (23.21%)	85.16 (23.55%)
DRA	66.37 (12.82%)	66.47 (12.84%)	66.75 (12.90%)	66.45 (12.84%)	66.89 (12.92%)	67.63 (13.07%)
FPK	79.40 (21.60%)	81.55 (22.19%)	81.09 (22.06%)	81.10 (22.06%)	80.39 (21.87%)	81.58 (22.19%)
GWN	91.89 (24.13%)	94.23 (24.74%)	94.39 (24.79%)	92.43 (24.27%)	95.07 (24.96%)	94.77 (24.88%)
PSU	98.24 (29.17%)	99.38 (29.51%)	99.92 (29.67%)	98.31 (29.19%)	100.40 (29.81%)	101.05 (30.00%)
SXF	84.27 (22.96%)	85.57 (23.32%)	86.84 (23.66%)	84.66 (23.07%)	85.90 (23.41%)	87.48 (23.84%)
TBL	105.31 (24.64%)	107.14 (25.07%)	106.89 (25.01%)	107.52 (25.16%)	108.71 (25.44%)	107.23 (25.09%)

7. Combining forecasts

The verification results thus far have been rather gloomy—hourly updated NWP forecasts, Kalman filtered or not, do not possess significant advantage over the arguably simpler time series methods, in particular, the optimal convex combination of climatology and persistence. There are, however, another "trick" which forecasters often use to improve performance, that is, combining forecasts. The idea of forecast combination, or ensemble forecasting, as referred to by weather forecasters, is a proven one, not theoretically, but pragmatically. By (weighted) averaging forecasts generated by different methods, it can be shown mathematically that the combined forecasts would always outperform the average member, see Appendix A of [59] for proof. Moreover, although there is no guarantee, one *very often* can receive a set of combined forecasts that is better than all members. Since the literature on forecast combination is bulky and rich, technical discussions are not reiterated. Instead, the readers are referred to some recent, and some not so recent but useful, reviews [60–65].

For the present purpose, 1-h-ahead forecasts from three time series methods (ARIMA, ETS, and REF) and two sets of filtered 1-h-ahead NWP forecasts (RAP+KF and HRRR+KF) are considered as member forecasts. A total of three combining strategies are used: (1) combining only time series forecasts, (2) combining only NWP forecasts, and (3) combining all five sets of forecasts; these strategies are denoted with "TS+Comb", "NWP+Comb", and "All+Comb", respectively. Simple averaging is used as the combination in all cases. The RMSE results are tabulated in Tables 6.

One must note the encouraging RMSE results in Table 6, in that, the combined forecasts dominate all component forecasts in all cases. Particularly noteworthy is the fact that combining all five component forecasts leads to best results at all SURFRAD stations. The amount of improvement from best component forecasts is sizeable. What should be also concluded is that combining both time series forecasts and NWP forecasts is more rewarding than combining either time series forecasts or NWP forecasts alone. The explanation is simple: time series forecasts and NWP forecasts are generated based on very distinctive

Table 6
1-h-ahead RMSE, in W/m^2 (%), of five sets of component forecasts and three sets of combined forecasts. TS + Comb combines ARIMA, ETS, and REF; NWP + Comb combines RAP + KF and HRRR + KF; and All + Comb combines all five component forecasts.

stn.	component forecasts					combined forecasts		
	ARIMA	ETS	REF	RAP + KF	HRRR + KF	TS + Comb	NWP + Comb	All + Comb
BON	87.71 (23.82%)	87.34 (23.72%)	85.98 (23.35%)	84.74 (23.02%)	83.39 (22.65%)	82.83 (22.50%)	82.69 (22.46%)	**79.43 (21.57%)**
DRA	69.13 (13.21%)	71.18 (13.60%)	66.81 (12.77%)	67.12 (12.83%)	67.23 (12.85%)	66.44 (12.70%)	66.65 (12.74%)	**64.14 (12.26%)**
FPK	82.42 (22.09%)	85.32 (22.87%)	80.58 (21.60%)	80.19 (21.49%)	81.85 (21.94%)	78.95 (21.16%)	79.67 (21.35%)	**76.41 (20.48%)**
GWN	95.97 (24.88%)	94.47 (24.49%)	92.29 (23.93%)	92.92 (24.09%)	93.48 (24.24%)	89.52 (23.21%)	91.69 (23.77%)	**86.91 (22.53%)**
PSU	99.41 (29.05%)	102.17 (29.86%)	96.85 (28.31%)	99.07 (28.96%)	99.14 (28.98%)	95.25 (27.84%)	97.56 (28.52%)	**93.09 (27.21%)**
SXF	84.79 (22.84%)	85.72 (23.09%)	83.58 (22.51%)	84.91 (22.87%)	85.36 (22.99%)	80.76 (21.75%)	84.02 (22.63%)	**79.23 (21.34%)**
TBL	106.54 (24.65%)	108.74 (25.16%)	105.46 (24.40%)	106.44 (24.63%)	108.75 (25.16%)	103.15 (23.86%)	105.52 (24.41%)	**100.50 (23.25%)**

mechanisms—the former ones are generated using data extrapolation, and the latter using dynamical weather modeling. As diversification is perhaps the most vital concern during forecast combination, these different forecast-generating mechanisms align well with such principle. In other words, a set of NWP forecast would be less correlated with a set of time series forecasts than another set of NWP forecasts; this is also true for time series forecasts. In fact, this has been reported by Yang and Dong [66] that by including just one additional NWP model into a time series ensemble, day-ahead solar forecasting performance can be improved substantially. At present, the same conclusion has been empirically reached, through an hourly solar forecasting case study.

8. Conclusion

This article has discussed several important aspects of using hourly updated numerical weather prediction (NWP) for solar forecasting. Throughout the article, ground-based research-grade measurements from the Surface Radiation Budget Network and NWP forecasts from the National Oceanic and Atmospheric Administration's Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) are taken as examples. Nevertheless, the present findings are believed to be transferable to circumstances with other NWP models. To make the conclusion concise, the findings are summarized in bullet points below.

- NWP forecast accuracy drops as forecast horizon increases. However, the rate of performance deterioration varies from location to location. The internal relative accuracy (IRA) computed between forecasts at the nearest and further horizons can be viewed as a proxy for predictability.
- Predictability is interpreted as the potential, or margin of improvements, to outperform a naïve reference method. Contrary to the conventional thinking, desert climates with predominant clear skies have a low predictability, because more elaborated models do not outperform naïve reference method at any significant rate.
- When optimal convex combination of climatology and persistence is used as the standard of reference, more elaborate models, such as the time series models or NWP models, often have negative forecast skill. This suggests that some superiority claims in the literature may have been exaggerating; the positive skills may be due to the poor choice of reference method, rather than the good performance of the model of interest.
- Kalman filtering is able to remove the inherent bias in NWP forecasts. Nonetheless, the superiority of filtered NWP forecasts over reference forecasts still cannot be guaranteed.
- Performing Kalman filtering on hourly updated NWP forecasts leads to only slightly better accuracies than performing Kalman filtering on 6-hourly updated or day-ahead NWP forecasts. In fact, the accuracy differences can be practically ignored.
- Averaging both time series forecasts and NWP forecasts is able to obtain best performance at all stations, which is otherwise unachievable if only time series or only NWP forecasts are combined. This finding confirms the necessity of ensemble forecasting (also known as multi-modeling).
- In order for the NWP forecasts to be most useful in an hourly solar forecasting setting, they must be post-processed first, and then combined with other methods of hour-ahead forecasting.

CRediT authorship contribution statement

Gang Zhang: Conceptualization, Methodology, Validation, Writing – original draft. **Dazhi Yang:** Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **George Galanis:** Methodology, Validation, Writing – review & editing. **Emmanouil Androulakis:** Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rser.2021.111768. This section contains the instructions for reproducing all results reported in the article. The code is written in R, and several packages (and, of course, their dependencies, see http://stat.ethz.ch/R-manual/R-patched/library/tools/html/package.dependencies.html) need to be installed before the scripts can be executed. The main EMOS routine only requires two packages:

- Package `dplyr` (https://cran.r-project.org/web/packages/dplyr/index.html) is a fast, consistent tool for working with dataframe-like objects, both in memory and out of memory.
- Package `lubridate` (https://cran.r-project.org/web/packages/lubridate/index.html) deals with date and time handling during R programming.
- Package `forecast` (https://cran.r-project.org/web/packages/forecast/index.html) is the famous time series forecasting package, which contains the routines for ARIMA and ETS, with automatic model-selection capability.

Other R packages that used for plotting and tabulating the results include `ggplot2`, `xables`, and `ggthemes`.

Data: A total 21 data files are included. These data files can be divided into three groups, each containing seven files, which correspond to seven SURFRAD stations. These files are:

- `xxx_2020_a.txt` (where "xxx" denotes the three-letter station abbreviation) contains the aggregated hourly ground-based measurements, hourly McClear clear-sky GHI, and hourly updated NWP forecasts. There are a total of 41 columns of forecasts, corresponding to 0–21-h-ahead RAP forecasts and 0–18-h-ahead HRRR forecasts, with column names `rap0`, ⋯, `rap21`, `hrrr0`, ⋯, `hrrr18`. In each row, the forecasts are made *for* the time stamp of the row. For example, the row "2020-01-01 16:00:00" contains 41 forecasts made for that time stamp, but issued at different times—e.g., the value under `rap0` refers to the 0-h-ahead RAP forecast (i.e., analysis) made at 2020-01-01 16:00:00, the value under `rap1` refers to the 1-h-ahead RAP forecast made at 2020-01-01 15:00:00, the value under `rap2` refers to the 2-h-ahead RAP forecast made at 2020-01-01 14:00:00, so on and so forth.

- `xxx_2020_b.txt` also contains the aggregated hourly ground-based measurements and McClear clear-sky GHI, but with hourly, 6-hourly, and daily updated NWP forecasts. There are a total of 6 columns of forecasts, with column names: `rap.u1` (hourly updated RAP forecasts), `rap.u6` (6-hourly updated RAP forecasts), `rap.u24` (daily updated RAP forecasts), `hrrr.u1` (hourly updated HRRR forecasts), `hrrr.u6` (6-hourly updated HRRR forecasts) and `hrrr.u24` (daily updated HRRR forecasts). The forecasts in each row are made *for* the time stamp of that row.
- `xxx.RData` contains the time series forecasts, generated with R script `gen.TS.fcst.R`, which is used for benchmarking.

Code: A total of seven R scripts are included. To use these scripts, the user only needs to change the working directory.

- `functions.R` contains the main routine for Kalman filtering, as well as some functions for error computation.
- `gen.TS.fcst.R` generates time series forecasts, using ARIMA, ETS and optimal convex climatology–persistence combination, and then saves the results in `RData` files. The description which corresponds to the code in this script is found in Section 5.2.
- `IRA.plot.R` reproduces Fig. 1, as shown in Section 5.1.
- `TS.skill.R` reproduces Fig. 2, as shown in Section 5.2.
- `Kalman.a.R` reproduces Tables 3 and 4, as shown in Section 6.2.
- `Kalman.b.R` reproduces Table 5, as shown in Section 6.3.
- `combine.R` reproduces Table 6, as shown in Section 7.

References

[1] Sengupta M, Habte A, Wilbert S, Gueymard C, Remund J. Best practices handbook for the collection and use of solar resource data for solar energy applications. Technical report NREL/TP-5D00-77635, Golden, CO (United States): National Renewable Energy Lab.(NREL); 2021.

[2] Holmgren WF, Hansen CW, Mikofski MA. Pvlib python: a python package for modeling solar energy systems. J Open Source Softw 2018;3(29):884. http://dx.doi.org/10.21105/joss.00884.

[3] Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. Appl Energy 2021;283:116239. http://dx.doi.org/10.1016/j.apenergy.2020.116239.

[4] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CFM, David M, Frimane Â, Gueymard CA, Hong T, Kay MJ, Killinger S, Kleissl J, Lauret P, Lorenz E, van der Meer D, Paulescu M, Perez R, Perpiñán-Lamigueiro O, Peters IM, Reikard G, Renné D, Saint-Drenan Y-M, Shuai Y, Urraca R, Verbois H, Vignola F, Voyant C, Zhang J. Verification of deterministic solar forecasts. Sol Energy 2020;210:20–37. http://dx.doi.org/10.1016/j.solener.2020.04.019, Special Issue on Grid Integration.

[5] Murphy AH. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecast 1993;8(2):281–93. http://dx.doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

[6] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101. http://dx.doi.org/10.1016/j.solener.2017.11.023, Advances in Solar Resource Assessment and Forecasting.

[7] van der Meer DW, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512. http://dx.doi.org/10.1016/j.rser.2017.05.212.

[8] Ganger D, Zhang J, Vittal V. Forecast-based anticipatory frequency control in power systems. IEEE Trans Power Syst 2018;33(1):1004–12. http://dx.doi.org/10.1109/TPWRS.2017.2705761.

[9] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. Int J Forecast 2016;32(3):914–38. http://dx.doi.org/10.1016/j.ijforecast.2015.11.011.

[10] Yang D, Quan H, Disfani VR, Liu L. Reconciling solar forecasts: Geographical hierarchy. Sol Energy 2017;146:276–86. http://dx.doi.org/10.1016/j.solener.2017.02.010.

[11] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39(6):535–76. http://dx.doi.org/10.1016/j.pecs.2013.06.002.

[12] Yang D, van der Meer D. Post-processing in solar forecasting: Ten overarching thinking tools. Renew Sustain Energy Rev 2021;140:110735. http://dx.doi.org/10.1016/j.rser.2021.110735.

[13] Yang D, Li W, Yagli GM, Srinivasan D. Operational solar forecasting for grid integration: Standards, challenges, and outlook. Sol Energy 2021;224:930–7. http://dx.doi.org/10.1016/j.solener.2021.04.002.

[14] Sweeney C, Bessa RJ, Browell J, Pinson P. The future of forecasting for renewable energy. WIREs Energy Environ 2020;9(2):e365. http://dx.doi.org/10.1002/wene.365.

[15] Pedro HTC, Larson DP, Coimbra CFM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. J Renew Sustain Energy 2019;11(3):036102. http://dx.doi.org/10.1063/1.5094494.

[16] Yang D. Validation of the 5-min irradiance from the National Solar Radiation Database (NSRDB). J Renew Sustain Energy 2021;13(1):016101. http://dx.doi.org/10.1063/5.0030992.

[17] Benjamin SG, Weygandt SS, Brown JM, Hu M, Alexander CR, Smirnova TG, Olson JB, James EP, Dowell DC, Grell GA, Lin H, Peckham SE, Smith TL, Moninger WR, Kenyon JS, Manikin GS. A North American hourly assimilation and model forecast cycle: The rapid refresh. Mon Weather Rev 2016;144(4):1669–94. http://dx.doi.org/10.1175/MWR-D-15-0242.1.

[18] Yang D, Wu E, Kleissl J. Operational solar forecasting for the real-time market. Int J Forecast 2019;35(4):1499–519. http://dx.doi.org/10.1016/j.ijforecast.2019.03.009.

[19] Athanasopoulos G, Hyndman RJ, Kourentzes N, Petropoulos F. Forecasting with temporal hierarchies. European J Oper Res 2017;262(1):60–74. http://dx.doi.org/10.1016/j.ejor.2017.02.046.

[20] Yang D, Quan H, Disfani VR, Rodríguez-Gallegos CD. Reconciling solar forecasts: Temporal hierarchy. Sol Energy 2017;158:332–46. http://dx.doi.org/10.1016/j.solener.2017.09.055.

[21] Bjerknes V. The problem of weather forecasting as a problem in mechanics and physics. In: Shapiro MA, Grønås S, editors. The Life Cycles of Extratropical Cyclones. Boston, MA: American Meteorological Society; 1999, p. 1–4.

[22] Richardson LF. Weather Prediction by Numerical Process. Cambridge University Press; 1922.

[23] Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. Nature 2015;525(7567):47–55. http://dx.doi.org/10.1038/nature14956.

[24] Benjamin SG, Dévényi D, Weygandt SS, Brundage KJ, Brown JM, Grell GA, Kim D, Schwartz BE, Smirnova TG, Smith TL, Manikin GS. An hourly assimilation–forecast cycle: The RUC. Mon Weather Rev 2004;132(2):495–518. http://dx.doi.org/10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2.

[25] Mass CF, Ovens D, Westrick K, Colle BA. Does increasing horizontal resolution produce more skillful forecasts?: The results of two years of real-time numerical weather prediction over the Pacific northwest. Bull Am Meteorol Soc 2002;83(3):407–30. http://dx.doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2.

[26] Mathiesen P, Kleissl J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Sol Energy 2011;85(5):967–77. http://dx.doi.org/10.1016/j.solener.2011.02.013.

[27] Zhang X, Li Y, Lu S, Hamann H, Hodge B, Lehman B. A solar time based analog ensemble method for regional solar power forecasting. IEEE Trans Sustain Energy 2019;10(1):268–79. http://dx.doi.org/10.1109/TSTE.2018.2832634.

[28] Yang D, Perez R. Can we gauge forecasts using satellite-derived solar irradiance? J Renew Sustain Energy 2019;11(2):023704. http://dx.doi.org/10.1063/1.5087588.

[29] Alexander C, Weygandt S, Benjamin S, Dowell D, Hu M, Smirnova T, Olson J, Kenyon J, Grell G, James E, Lin H, Ladwig T, Brown J, Alcott T, Jankov I. WRF-Arw research to operations update: The rapid-refresh (RAP) version 4, High-Resolution Rapid Refresh (HRRR). 2017.

[30] Diebold FX, Kilian L. Measuring predictability: theory and macroeconomic applications. J Appl Econometrics 2001;16(6):657–69. http://dx.doi.org/10.1002/jae.619.

[31] Gneiting T. Making and evaluating point forecasts. J Amer Statist Assoc 2011;106(494):746–62. http://dx.doi.org/10.2307/41416407, URL: http://www.jstor.org/stable/41416407.

[32] Yang X, Yang D, Bright JM, Yagli GM, Wang P. On predictability of solar irradiance. J Renew Sustain Energy 2021;13(5):056501. http://dx.doi.org/10.1063/5.0056918.

[33] Murphy AH. Forecast verification: Its complexity and dimensionality. Mon Weather Rev 1991;119(7):1590–601. http://dx.doi.org/10.1175/1520-0493(1991)119<1590:FVICAD>2.0.CO;2.

[34] Yang D. Making reference solar forecasts with climatology, persistence, and their optimal convex combination. Sol Energy 2019;193:981–5. http://dx.doi.org/10.1016/j.solener.2019.10.006.

[35] Yang D. Standard of reference in operational day-ahead deterministic solar forecasting. J Renew Sustain Energy 2019;11(5):053702. http://dx.doi.org/10.1063/1.5114985.

[36] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98. http://dx.doi.org/10.1016/j.rser.2019.02.006.

[37] Yagli GM, Yang D, Gandhi O, Srinivasan D. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl Energy 2020;259:114122. http://dx.doi.org/10.1016/j.apenergy.2019.114122.

[38] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J Sel Top Appl Earth Obs Remote Sens 2009;2(1):2–10. http://dx.doi.org/10.1109/JSTARS.2009.2020300.

[39] Mejia JF, Giordano M, Wilcox E. Conditional summertime day-ahead solar irradiance forecast. Sol Energy 2018;163:610–22. http://dx.doi.org/10.1016/j.solener.2018.01.094.
[40] Yang D. On post-processing day-ahead NWP forecasts using Kalman filtering. Sol Energy 2019;182:179–81. http://dx.doi.org/10.1016/j.solener.2019.02.044.
[41] Pelland S, Galanis G, Kallos G. Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. Prog Photovolt, Res Appl 2013;21(3):284–96. http://dx.doi.org/10.1002/pip.1180.
[42] Diagne M, David M, Boland J, Schmutz N, Lauret P. Post-processing of solar irradiance forecasts from WRF model at Reunion Island. Sol Energy 2014;105:99–108. http://dx.doi.org/10.1016/j.solener.2014.03.016.
[43] Yang D. SolarData: An R package for easy access of publicly available solar datasets. Sol Energy 2018;171:A3–12. http://dx.doi.org/10.1016/j.solener.2018.06.107.
[44] Yang D. SolarData package update v1.1: R functions for easy access of Baseline Surface Radiation Network (BSRN). Sol Energy 2019;188:970–5. http://dx.doi.org/10.1016/j.solener.2019.05.068.
[45] Yang D. A correct validation of the National Solar Radiation Data Base (NSRDB). Renew Sustain Energy Rev 2018;97:152–5. http://dx.doi.org/10.1016/j.rser.2018.08.023.
[46] Bowman DC, Lees JM. Near real time weather and ocean model data access with rNOMADS. Comput Geosci 2015;78:88–95. http://dx.doi.org/10.1016/j.cageo.2015.02.013.
[47] Gschwind B, Wald L, Blanc P, Lefèvre M, Schroedter-Homscheidt M, Arola A. Improving the McClear model estimating the downwelling solar radiation at ground level in cloud-free conditions – McClear-v3. Meteorol Z 2019;28(2):147–63. http://dx.doi.org/10.1127/metz/2019/0946.
[48] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, Wald L, Schroedter-Homscheidt M, Hoyer-Klick C, Arola A, Benedetti A, Kaiser JW, Morcrette J-J. Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos Meas Tech 2013;6(9):2403–18. http://dx.doi.org/10.5194/amt-6-2403-2013.
[49] Yang D. Choice of clear-sky model in solar forecasting. J Renew Sustain Energy 2020;12(2):026101. http://dx.doi.org/10.1063/5.0003495.
[50] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F. forecast: Forecasting functions for time series and linear models. R package version 8.13, 2020, URL: https://pkg.robjhyndman.com/forecast/.
[51] Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for r. J Stat Softw 2008;26(3):1–22. http://dx.doi.org/10.18637/jss.v027.i03.
[52] Yang D, Jirutitijaroen P, Walsh WM. Hourly solar irradiance time series forecasting using cloud cover index. Sol Energy 2012;86(12):3531–43. http://dx.doi.org/10.1016/j.solener.2012.07.029.
[53] Yang D, Sharma V, Ye Z, Lim LI, Zhao L, Aryaputera AW. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. Energy 2015;81:111–9. http://dx.doi.org/10.1016/j.energy.2014.11.082.
[54] Lauret P, Voyant C, Soubdhan T, David M, Poggi P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Sol Energy 2015;112:446–57. http://dx.doi.org/10.1016/j.solener.2014.12.014, URL: https://www.sciencedirect.com/science/article/pii/S0038092X14006057.
[55] Dong Z, Yang D, Reindl T, Walsh WM. Short-term solar irradiance forecasting using exponential smoothing state space model. Energy 2013;55:1104–13. http://dx.doi.org/10.1016/j.energy.2013.04.027.
[56] Pedro HTC, Coimbra CFM, David M, Lauret P. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. Renew Energy 2018;123:191–203. http://dx.doi.org/10.1016/j.renene.2018.02.006.
[57] Murphy AH. Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather Forecast 1992;7(4):692–8. http://dx.doi.org/10.1175/1520-0434(1992)007<0692:CPATLC>2.0.CO;2.
[58] Galanis G, Louka P, Katsafados P, Pytharoulis I, Kallos G. Applications of Kalman filters based on non-linear functions to numerical weather predictions. Ann Geophys 2006;24(10):2451–60. http://dx.doi.org/10.5194/angeo-24-2451-2006.
[59] Yang D. Ensemble model output statistics as a probabilistic site-adaptation tool for satellite-derived and reanalysis solar irradiance. J Renew Sustain Energy 2020;12(1):016102. http://dx.doi.org/10.1063/1.5134731.
[60] Thomson ME, Pollock AC, Önkal D, Gönül MS. Combining forecasts: Performance and coherence. Int J Forecast 2019;35(2):474–84. http://dx.doi.org/10.1016/j.ijforecast.2018.10.006.
[61] Wallis KF. Combining forecasts – Forty years later. Appl Financial Econ 2011;21(1–2):33–41. http://dx.doi.org/10.1080/09603107.2011.523179.
[62] Hibon M, Evgeniou T. To combine or not to combine: selecting among forecasts and their combinations. Int J Forecast 2005;21(1):15–24. http://dx.doi.org/10.1016/j.ijforecast.2004.05.002.
[63] Armstrong JS. Combining forecasts. In: Principles of Forecasting. Springer; 2001, p. 417–39.
[64] Diebold FX, Lopez JA. Forecast evaluation and combination. In: Handbook of Statistics, Vol. 14. 1996, p. 241–68. http://dx.doi.org/10.1016/S0169-7161(96)14010-4.
[65] Clemen RT. Combining forecasts: A review and annotated bibliography. Int J Forecast 1989;5(4):559–83. http://dx.doi.org/10.1016/0169-2070(89)90012-5.
[66] Yang D, Dong Z. Operational photovoltaics power forecasting using seasonal time series ensemble. Sol Energy 2018;166:529–41. http://dx.doi.org/10.1016/j.solener.2018.02.011.